

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2004年6月3日 (03.06.2004)

PCT

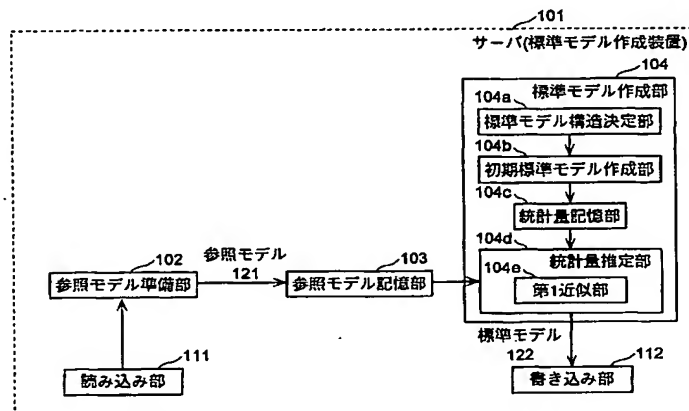
(10) 国際公開番号
WO 2004/047076 A1

- (31) 国際特許分類: G10L 15/06, G06K 9/68
(32) 出願番号: PCT/JP2003/014626
(33) 出願日: 2003年11月18日 (18.11.2003)
(25) 国際出願の言語: 日本語
(26) 国際公開の言語: 日本語
(30) 優先権データ:
特願 2002-338652
2002年11月21日 (21.11.2002) JP
特願 2003-89179 2003年3月27日 (27.03.2003) JP
特願 2003-284489 2003年7月31日 (31.07.2003) JP
(71) 出願人 (米国を除く全ての指定国について): 松下電器産業株式会社 (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) [JP/JP]; 〒571-8501 大阪府門真市大字門真 1006 番地 Osaka (JP).
(72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 芳澤 伸一 (YOSHIZAWA, Shinichi) [JP/JP]; 〒573-0064 大阪府枚方市北中振2丁目3番32-304号 Osaka (JP).
(74) 代理人: 新居 広守 (NII, Hiromori); 〒532-0011 大阪府大阪市淀川区西中島3丁目11番26号 新大阪末広センタービル3F 新居国際特許事務所内 Osaka (JP).
(81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[続葉有]

(54) Title: STANDARD MODEL CREATING DEVICE AND STANDARD MODEL CREATING METHOD

(54) 発明の名称: 標準モデル作成装置及び標準モデル作成方法



102...REFERENCE MODEL PREPARING UNIT
111...READING UNIT
121...REFERENCE MODEL
103...REFERENCE MODEL STORAGE UNIT
101...SERVER (STANDARD MODEL CREATING DEVICE)
104...STANDARD MODEL CREATING UNIT
104a...STANDARD MODEL STRUCTURE DETERMINING SECTION
104b...INITIAL STANDARD MODEL PREPARING SECTION
104c...STATISTIC QUANTITY STORAGE SECTION
104d...STATISTIC QUANTITY ESTIMATING SECTION
104e...FIRST APPROXIMATING SECTION
122...STANDARD MODEL
112...WRITING MODEL

(57) Abstract: A standard model creating device for providing a high-precision standard model used for pattern recognition such as speech recognition, character recognition, or image recognition by using a probability model based on a hidden Markov model, the Bayesian theory, linear discrimination analysis, intention interpretation using a probability model such as a Bayesian net, and data-mining performed by using a probability model. The standard model creating device comprises a reference model preparing unit (102) for preparing

[続葉有]



(84) 指定国 (広域): ARIPO 特許 (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

one or more reference models, a reference model storage unit (103) storing a reference model (121) prepared by the reference model preparing unit (102), and a standard model creating unit (104) for creating a standard model (122) by calculating a statistic quantity of the standard model in such a way that the probability or likelihood of one or more reference models stored in the reference model storage unit (103) is maximized or locally maximized.

(57) 要約: 隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解、確率モデルによるデータマイニングなどに用いる高精度な標準モデルを提供する標準モデル作成装置であって、1以上の参照モデルを準備する参照モデル準備部(102)と、参照モデル準備部(102)が準備した参照モデル(121)を記憶する参照モデル記憶部(103)と、参照モデル記憶部(103)が記憶している1以上の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデル(122)を作成する標準モデル作成部(104)とを備える。

明 細 書

標準モデル作成装置及び標準モデル作成方法

技術分野

- 5 本発明は、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、確率モデルによる人物検出、指紋認証、顔認証、虹彩認証（対象を認識して特定の対象かどうかを判断する）、株価予測、天気予測などの予測（状況を認識して判断する）、複数の話者音声の合成、複数の顔画像などの合成（合成したモデルを人が認識して楽しむ）などに用いられる標準モデルの作成装置及びその方法に関する。
- 10

15 背景技術

- 近年、インターネットなどの普及により、ネットワークの大容量化、通信コストの低価格化が進んでいる。このため、ネットワークを利用することで、多くの認識用モデル（参照モデル）を収集することが可能となってきた。例えば、音声認識において、様々な研究機関で配布している多くの音声認識用モデル（子供用モデル、成人用モデル、高齢者
- 20 用モデル、自動車内用モデル、携帯電話用モデルなど）をインターネットによりダウンロードすることが可能となってきた。また、ネットワークによる機器連携により、カーナビゲーションシステムなどで利用する音声認識用モデルをテレビやパソコンなどにダウンロードできるようになってきている。また、意図理解において、各地の様々な人の経験を学習した認識用モデルを、ネットワークを通して収集することが可能
- 25

となってきた。

また、認識技術の発展により、認識用モデルは、パソコン、テレビのリモコン、携帯電話、カーナビゲーションシステムなど、CPUパワー、メモリ量などの仕様の異なる幅広い機器に利用されるようになってきている。また、セキュリティなどの認識精度が要求されるアプリケーションや、テレビのリモコンでの操作のように認識結果が出力されるまでの時間の速さが要求されるアプリケーションなど、要求仕様の異なる幅広いアプリケーションに利用されるようになってきている。

また、認識技術は、認識対象の異なる多くの環境で利用されるようになってきている。例えば、音声認識において、子供の声、成人の声、高齢者の声を認識したり、自動車内での声、携帯電話での声を認識するなど、多くの環境で利用される。

これらの社会環境の変化を鑑みると、多くの認識用モデル（参照モデル）を有効に活用することで、機器やアプリケーションの仕様、利用環境に適した精度の高い認識用モデル（標準モデル）を短時間に作成して利用者に提供することが望まれると考えられる。

音声認識などのパターン認識の分野では、認識用の標準モデルとして確率モデルを用いる方法が近年注目されており、特に、隠れマルコフモデル（以下HMMと呼ぶ）や混合ガウス分布モデル（以下GMMと呼ぶ）が広く用いられている。また、意図理解において、意図、知識、嗜好などを表す標準モデルとして確率モデルを用いる方法が近年注目されており、特に、ベイジアンネットなどが広く用いられている。また、データマイニングの分野で、データを分類するために各カテゴリの代表モデルとして確率モデルを用いる方法が注目されており、GMMなどが広く用いられている。また、音声認証、指紋認証、顔認証、虹彩認証などの認証の分野で、認証用の標準モデルとして確率モデルを用いる方法が注目

されており、GMMなどが用いられている。HMMにより表現される標準モデルの学習アルゴリズムとしてバウム・ウェルチ（Baum-Welch）の再推定の方法が広く用いられている（例えば、今井聖著、"音声認識"、pp.150-152、共立出版株式会社、1995年11月25日発行参照）。また、GMMにより表現される標準モデルの学習アルゴリズムとしてEM（Expectation-Maximization）アルゴリズムが広く用いられている（例えば、古井貞▲ひろ▼著、"音声情報処理"、pp.100-104、森北出版株式会社、1998年6月30日発行参照）。EMアルゴリズムでは、標準モデル

10 （式1）

$$\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)$$

（ここで、

（式2）

$$f(x; \mu_{f(m)}, \sigma_{f(m)}^2) \quad (m=1, 2, \dots, M_f)$$

15 はガウス分布を表し、

（式3）

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

はJ（ ≥ 1 ）次元の入力データを表す）における統計量である混合重み係数

20 （式4）

、 J (≥ 1) 次元の平均値

(式 5)

$$\mu_{f(m)} = (\mu_{f(m,1)}, \mu_{f(m,2)}, \dots, \mu_{f(m,J)}) \in R^J$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

及び J (≥ 1) 次元の分散値 (共分散行列の J 個の対角成分)

5 (式 6)

$$\sigma_{f(m)}^2 = (\sigma_{f(m,1)}^2, \sigma_{f(m,2)}^2, \dots, \sigma_{f(m,J)}^2) \in R^J$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

を、

N 個の学習データ

(式 7)

10 $x[i] = (x_{(1)}[i], x_{(2)}[i], \dots, x_{(J)}[i]) \in R^J \quad (i = 1, 2, \dots, N)$

を用いて、学習データに対する尤度

(式 8)

$$\log P = \sum_{i=1}^N \log \left[\sum_{m=1}^{M_f} \omega_{f(m)} f(x[i]; \mu_{f(m)}, \sigma_{f(m)}^2) \right]$$

を最大化もしくは極大化するように、

15 (式 9)

$$\omega_{f(m)} = \frac{\sum_{i=1}^N \gamma(x[i], m)}{\sum_{k=1}^{M_f} \sum_{i=1}^N \gamma(x[i], k)}$$

$$(m = 1, 2, \dots, M_f)$$

(式 10)

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^N \gamma(x[i], m) x_{(j)}}{\sum_{i=1}^N \gamma(x[i], m)}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

(式 11)

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^N \gamma(x[i], m) (x_{(j)} - \mu_{f(m,j)})^2}{\sum_{i=1}^N \gamma(x[i], m)}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

5

(ここで、

(式 12)

$$\gamma(x[i], m) = \frac{\omega_{f(m)} f(x[i]; \mu_{f(m)}, \sigma_{f(m)}^2)}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x[i]; \mu_{f(k)}, \sigma_{f(k)}^2)} \quad (m = 1, 2, \dots, M_f)$$

である) を利用して 1 以上繰り返して計算して学習を行う。また、ベ
 10 イズ推定法(例えば、繁樹算男著、"ベイズ統計入門"、p p. 42-53、
 東京大学出版会、1985年4月30日発行参照)などの方法も提案さ

れている。バウム・ウェルチの再推定の方法、EMアルゴリズム、ベイズ推定法のいずれの学習アルゴリズムも、学習データに対する確率（尤度）を最大化もしくは極大化するように標準モデルのパラメータ（統計量）を計算して標準モデルを作成する。これらの学習方法では、確率（尤度）を最大化もしくは極大化するという数学的な最適化が実現されている。

上記の学習方法を音声認識の標準モデルの作成に用いた場合、多様な話者や雑音などの音響的特徴量の変動に対応するために多数の音声データで標準モデルを学習することが望ましい。また、意図理解に用いた場合、多様な話者や状況などの変動に対応するために多数のデータで標準モデルを学習することが望ましい。また、虹彩認証に用いた場合、太陽光、カメラ位置・回転などの変動に対応するために多数の虹彩画像データで標準モデルを学習することが望ましい。しかしながら、このような多量のデータを取り扱う場合、学習に膨大な時間がかかるため、利用者に標準モデルを短時間に提供できない。また、多量のデータを蓄積するためのコストが膨大となる。また、ネットワークを利用してデータを収集した場合、通信コストが膨大となる。

一方、複数のモデル（以下、標準モデルの作成のために参照用として準備されるモデルを「参照モデル」と呼ぶ。）を合成することで標準モデルを作成する方法が提案されている。参照モデルは、多くの学習データを確率分布の母数（平均、分散など）で表現した確率分布モデルであり、多くの学習データの特徴を少数のパラメータ（母数）で集約したものである。以下に示す従来技術では、モデルはガウス分布で表現されている。

第1の従来方法では、参照モデルはGMMで表現されており、複数の参照モデルのGMMを重み付きで合成することで標準モデルを作成している（例えば、特開平4-125599号公報に開示された技術）。

また、第２の従来方法では、第１の従来方式に加えて、学習データに対する確率（尤度）を最大化あるいは極大化して線形結合された混合重みを学習することで標準モデルを作成している（例えば、特開平１０－２６８８９３号公報に開示された技術）。

- ５ また、第３の従来方法では、標準モデルの平均値を参照モデルの平均値の線形結合で表現し、入力データに対する確率（尤度）を最大化あるいは極大化して線形結合係数を学習することで標準モデルを作成している。ここでは学習データとして特定話者の音声データを用いており標準モデルを音声認識用の話者適応モデルとして用いている（例えば、M. J. F. Gales, "Cluster Adaptive Training For Speech Recognition", 1998年、ICSLP98予稿集、pp. 1783－1786）。
- 10

- また、第４の従来方法では、参照モデルは単一ガウス分布で表現されており、複数の参照モデルのガウス分布を合成したのちに、クラスタリングにより同一クラスに属するガウス分布を統合することで標準モデル
- 15
- を作成している（例えば、特開平９－８１１７８号公報に開示された技術）。

- また、第５の従来方法では、複数の参照モデルは同数の混合数の混合ガウス分布で表現され、各ガウス分布には１対１に対応した通し番号が付与されている。標準モデルは、同一の通し番号をもつガウス分布を合成することにより作成される。合成する複数の参照モデルは利用者に音響的に近い話者で作成されたモデルであり、作成させる標準モデルは話者適応モデルである（例えば、芳澤、外６名、"十分統計量と話者距離を用いた音韻モデルの教師なし学習法"、２００２年３月１日、電子情報通
- 20
- 信学会、Vol. J85-D-II、No. 3、pp. 382－389）。
- 25

しかしながら、第１の従来方法では、合成する参照モデル数の増加と

ともに標準モデルの混合数が増加して、標準モデルのための記憶容量、認識処理量が膨大となり実用的でない。また、仕様に依じて標準モデルの混合数を制御することができない。この課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

- 5 第2の従来方法では、合成する参照モデル数の増加とともに標準モデルの混合数が増加して、標準モデルのための記憶容量、認識処理量が膨大となり実用的でない。また、仕様に依じて標準モデルの混合数を制御することができない。また、標準モデルは、参照モデルの単純な混合和であり学習するパラメータが混合重みに限定されているため、高精度の
- 10 標準モデルが作成できない。また、標準モデルの作成において、多くの学習データを用いて学習を行っているため学習時間がかかる。これらの課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

- 第3の従来方法では、学習するパラメータが参照モデルの平均値の線形結合係数に限定されているため高精度の標準モデルが作成できない。
- 15 また、標準モデルの作成において、多くの学習データを用いて学習を行っているため学習時間がかかる。

- 第4の従来方法では、クラスタリングをヒューリスティックに行うため高精度の標準モデルを作成することが困難である。また、参照モデル
- 20 は単一のガウス分布であるため精度が低く、それらを統合した標準モデルの精度は低い。認識精度に関する課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

- 第5の従来方法では、標準モデルは、同一の通し番号をもつガウス分布を合成することにより作成されるが、最適な標準モデルを作成するためには、一般的には合成するガウス分布は1対1に対応するとは限らないため、認識精度が低下する。また、複数の参照モデルが異なる混合数
- 25

をもつ場合に標準モデルを作成することができない。また、一般的には、参照モデルにおけるガウス分布に通し番号が付与されておらず、この場合に標準モデルを作成することができない。また、仕様に依じて標準モデルの混合数を制御することができない。

5

発明の開示

そこで、本発明は、このような問題点に鑑みてなされたものであり、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットワークなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、株価予測、天気予測などの予測（状況を認識して判断する）などに用いられる高精度な標準モデルを作成する標準モデル作成装置等を提供することを目的とする。

また、本発明は、学習のためのデータや教師データを必要とすることなく、簡易に標準モデルを作成することが可能な標準モデル作成装置等を提供することをも目的とする。

さらに、本発明は、標準モデルを利用する認識の対象にふさわしい標準モデルを作成したり、標準モデルを用いて認識処理を実行する装置の仕様や環境に適した標準モデルを作成することが可能な汎用性及び柔軟性に優れた標準モデル作成装置等を提供することをも目的とする。

本発明で用いる「認識」とは、音声認識などの狭義の意味での認識だけではなく、パターンマッチング、識別、認証、ベイズ推定や予測など、確率で表現された標準モデルを利用するもの全般を意味する。

上記目的を達成するために、本発明に係る標準モデル作成装置は、事象の集合と事象または事象間の遷移の出力確率とによって定義される認識用のモデルである標準モデルを作成する装置であって、特定の対象を

認識するために予め作成されたモデルである 1 以上の参照モデルを記憶する参照モデル記憶手段と、前記参照モデル記憶手段に記憶された 1 以上の参照モデルに対する標準モデルの確率または尤度を最大化または極大化するように当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備えることを特徴とする。

たとえば、音声認識用の標準モデル作成装置として、音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである 1 以上の参照モデルを記憶する参照モデル記憶手段と、前記参照モデル記憶手段に格納された 1 以上の参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備え、前記標準モデル作成手段は、作成する標準モデルの構造を決定する標準モデル構造決定部と、構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成部と、初期値が決定された標準モデルの前記参照モデルに対する確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有することを特徴とする。

これによって、1 以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するように標準モデルの統計量が計算され、標準モデルが作成されるので、音声データ等の学習データや教師データを必要とすることなく簡易に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した高精度な標準モデルが作成される。

ここで、前記標準モデル作成装置は、さらに、外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、参照モデルを

作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段を備えてもよい。例えば、音声認識用に適用した場合であれば、音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである 1 以上の参照モデルを記憶するための参照モデル記憶手段と、外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、新たな参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段と、所定の構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル記憶手段に格納された 1 以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するように、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備えることを特徴とする。

これによって、標準モデル作成装置の外部から新たな参照モデルを取り込み、取り込んだ参照モデルに基づいた標準モデルの作成が可能となるので、様々な認識対象に対応した汎用性の高い標準モデル作成装置が実現される。

また、前記標準モデル作成装置は、さらに、認識の対象に関する情報である利用情報を作成する利用情報作成手段と、作成された前記利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から 1 以上の参照モデルを選択する参照モデル選択手段とを備え、前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、利用者の特徴、利用者の年齢、性別、利用環境などの

利用情報に基づいて、準備された複数の参照モデルの中から認識対象に適した参照モデルだけが選択され、それら参照モデルを統合した標準モデルが作成されるので、認識対象により特化した精度の高い標準モデルが作成される。

- 5 ここで、前記標準モデル作成装置は、さらに、前記利用情報と選択された参照モデルに関する情報との類似度を算出して、前記類似度が所定のしきい値以上であるか否かを判定して判定信号を作成する類似度判定手段を備えてもよい。

- 10 これによって、利用情報にふさわしい（近い）参照モデルが参照モデル記憶手段に存在しない場合に、参照モデルの準備の要求を行うことができる。

- 15 また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、認識の対象に関する情報である利用情報を前記端末装置から受信する利用情報受信手段と、受信された前記利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から1以上の参照モデルを選択する参照モデル選択手段とを備え、前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

- 20 これによって、通信路を介して送信されてきた利用情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

- 25 また、前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報を作成する仕様情報作成手段を備え、前記標準モデル作成手段は、前記仕様情報作成手段が作成した仕様情報に

基づいて、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、標準モデルを使用する装置のCPUパワー、記憶容量、要求される認識精度、要求される認識処理時間などの仕様情報に基づいて標準モデルが作成されるので、特定の仕様条件を満たす標準モデルの生成が可能となり、計算エンジン等の認識処理に必要なリソース環境に適した標準モデルの生成が実現される。

ここで、前記仕様情報は、例えば、標準モデルを使用するアプリケーションプログラムの種類に対応づけられた仕様を示すような情報であってもよい。そして、前記標準モデル作成装置は、さらに、標準モデルを使用するアプリケーションプログラムと標準モデルの仕様との対応を示すアプリケーション仕様対応データベースを前記仕様情報として保持する仕様情報保持手段を備え、前記標準モデル作成手段は、前記仕様情報保持手段に保持されたアプリケーション仕様対応データベースから、起動されるアプリケーションプログラムに対応する仕様を読み出し、読み出した仕様に基づいて、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、各アプリケーションごとに対応づけられた仕様に沿って標準モデルが作成されるので、アプリケーションごとに最適な標準モデルが作成され、標準モデルが使用される認識システム等における認識精度が向上される。

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報を前記端末装置から受信する仕様情報受信手段を備え、前記標準モデル作成手段は、前記仕様情報受信手段が受信

した仕様情報に基づいて、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

5 これによって、通信路を介して送信されてきた仕様情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

10 たとえば、前記参照モデル及び前記標準モデルは、1以上のガウス分布を用いて表現され、前記標準モデル作成手段は、前記仕様情報に基づいて、前記標準モデルの混合分布数（ガウス分布の数）を決定してもよい。

15 これによって、作成される標準モデルに含まれるガウス分布の混合分布数が動的に決定されることとなり、認識処理が実行される環境や要求仕様等に応じて標準モデルの構造を制御することが可能となる。例として、標準モデルを使用する認識装置のCPUパワーが小さい場合、記憶容量が小さい場合、要求される認識処理時間が短い場合などは標準モデルの混合分布数を少なく設定して仕様に合わせることができ、一方、要求される認識精度が高い場合などは混合分布数を多く設定して認識精度を高くすることができる。

20 なお、上記利用情報あるいは仕様情報を用いて標準モデルを作成する場合において、参照モデル準備手段は必ずしも必要ではない。たとえば、利用者の要求に基づいて、あるいは、利用者の要求とは無関係に、予め参照モデルを標準モデル作成装置内に記憶させた状態で標準モデル作成装置を出荷し、利用情報や仕様情報を用いて標準モデルを作成することが可能だからである。

25

また、前記参照モデル及び前記標準モデルは、1以上のガウス分布を

用いて表現され、前記参照モデル記憶手段は、少なくとも 1 対の参照モデルの混合分布数（ガウス分布の数）が異なる参照モデルを記憶し、前記標準モデル作成手段は、少なくとも 1 対の参照モデルの混合分布数（ガウス分布の数）が異なる参照モデルに対する前記標準モデルの確率又は
5 尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、混合分布数が異なる参照モデルに基づいて標準モデルが作成されるので、予め準備された多種多様な構造の参照モデルに基づく標準モデルの作成が可能となり、より認識対象に適した精度の高い標準モデルの作成が実現される。
10

また、前記標準モデル作成装置は、さらに、前記標準モデル作成手段が作成した標準モデルを記憶する標準モデル記憶手段を備えてもよい。

これによって、作成された標準モデルを一時的にバッファリングしておき、送信要求に対してすぐに出力したり、他の装置に提供するデータサーバとしての役割を果たしたりすることが可能となる。
15

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記標準モデル作成手段が作成した標準モデルを前記端末装置に送信する標準モデル送信手段を備えてもよい。

20 これによって、作成された標準モデルは空間的に離れた場所に置かれた外部装置に送信されるので、本標準モデル作成装置を標準モデル作成エンジンとして独立させたり、標準モデル作成装置を通信システムにおけるサーバとして機能させたりすることが可能になる。

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記端末装置から送信される参照モデルを受信する参照モデル受信手段を備え、前記標準モデル作
25

成手段は、少なくとも前記参照モデル受信手段が受信した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

5 これによって、端末装置が保持した利用環境にふさわしい参照モデルを、通信路を介して送信して、送信した参照モデルを用いて標準モデルを作成できるため、より認識対象に適した精度の高い標準モデルの作成が実現される。例として、利用者Aが環境Aで利用していた参照モデルAが端末装置に保持されており利用者Aは環境Bで利用したい場合、参照モデルAを利用して標準モデルを作成することにより、利用者Aの特徴を反映した精度の高い標準モデルを作成することができる。

10 また、前記参照モデル準備手段は、さらに、前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行ってもよい。たとえば、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記端末装置から送信される参照モデルを受信する参照モデル受信手段を備え、前記参照モデル準備手段は、前記参照モデル受信手段が受信した参照モデルを用いて前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行ってもよい。

20 これによって、準備される参照モデルの追加、更新等が行われるので、様々な認識対象用のモデルを参照モデルとして追加したり、より精度の高い参照モデルに置き換えたりすることが可能となり、更新した参照モデルによる標準モデルの再生成や、生成された標準モデルを参照モデルとして再び標準モデルを作成するというフィードバックによる学習等が可能となる。

25 また、前記標準モデル作成手段は、作成する標準モデルの構造を決定する標準モデル構造決定部と、構造が決定された前記標準モデルを特定

する統計量の初期値を決定する初期標準モデル作成部と、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有するように構成してもよい。このとき、前記初期標準モデル作成部は、前記統計量推定部が標準モデルの統計量を計算するために用いる、1以上の前記参照モデルを用いて前記標準モデルを特定する統計量の初期値を決定してもよい。たとえば、前記初期標準モデル作成部は、標準モデルの種類を識別するクラスIDに基づいて、前記初期値を決定してもよい。具体的には、前記初期標準モデル作成部は、前記クラスIDと前記初期値と前記参照モデルとの対応を示す対応表を保持し、前記対応表に従って、前記初期値を決定してもよい。

これによって、標準モデルが使用される認識の対象の種類ごとにクラスIDを付与しておくことで、最終的に必要とされる標準モデルと共通の性質をもつ初期標準モデルを使用することができるので、精度の高い標準モデルが作成される。

以上のように、本発明により、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解(意図の認識)、確率モデルによるデータマイニング(データ特性の認識)、確率モデルによる人物検出、指紋認証、顔認証、虹彩認証(対象を認識して特定の対象かどうかを判断する)、株価予測、天気予測などの予測(状況を認識して判断する)などに用いる高精度な標準モデルが提供され、その実用的価値は極めて高い。

なお、本発明は、このような標準モデル作成装置として実現することができるだけでなく、標準モデル作成装置が備える特徴的な構成要素をステップとする標準モデル作成方法として実現したり、それらのステッ

プをコンピュータに実行させるプログラムとして実現したりすることができる。そして、そのプログラムをCD-ROM等の記録媒体やインターネット等の伝送媒体を介して配信することができるのは言うまでもない。

5

図面の簡単な説明

図1は、本発明の第1の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図2は、同サーバの動作手順を示すフローチャートである。

10 図3は、図1における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

図4は、図2におけるステップS101（標準モデルの作成）の詳細な手順を示すフローチャートである。

15 図5は、図1における第1近似部104eによる近似計算を説明する図である。

図6は、参照モデルを選択する際の画面表示例を示す図である。

図7(a)は、作成する標準モデルの構造（混合分布数）を指定する際の画面表示例を示し、図7(b)は、仕様情報を選択する際の画面表示例を示す図である。

20 図8は、標準モデルを作成しているときの進捗状況を示す画面表示例を示す図である。

図9は、本発明の第2の実施の形態における標準モデル作成装置に係るSTBの全体構成を示すブロック図である。

図10は、同STBの動作手順を示すフローチャートである。

25 図11は、図10における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

図 1 2 は、図 1 0 における第 2 近似部による近似計算を説明する図である。

図 1 3 は、本発明の第 3 の実施の形態における標準モデル作成装置に係る P D A の全体構成を示すブロック図である。

5 図 1 4 は、同 P D A の動作手順を示すフローチャートである。

図 1 5 は、図 1 3 における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

図 1 6 は、同 P D A の選択画面の一例を示す。

10 図 1 7 は、図 1 3 における統計量推定部による統計量の推定手順を示す概念図である。

図 1 8 は、図 1 3 における第 3 近似部による近似計算を説明する図である。

図 1 9 は、本発明の第 4 の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

15 図 2 0 は、同サーバの動作手順を示すフローチャートである。

図 2 1 は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図 2 2 は、利用情報としての個人情報を入力する際の画面表示例を示す図である。

20 図 2 3 は、本発明の第 5 の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図 2 4 は、同サーバの動作手順を示すフローチャートである。

図 2 5 は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

25 図 2 6 は、本発明の第 6 の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図 2 7 は、同サーバの動作手順を示すフローチャートである。

図 2 8 は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図 2 9 は、本発明の第 7 の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図 3 0 は、同サーバの動作手順を示すフローチャートである。

図 3 1 は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図 3 2 は、本発明の第 8 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。

図 3 3 は、携帯電話 9 0 1 の動作手順を示すフローチャートである。

図 3 4 は、参照モデル記憶部に格納されている参照モデルの一例を示す図である。

図 3 5 は、新たに参照モデル記憶部に格納された参照モデルの一例を示す図である。

図 3 6 は、利用情報を作成するときの画面表示例を示す図である。

図 3 7 は、参照モデルを準備するときの画面表示例を示す図である。

図 3 8 は、第 3 近似部を用いて作成した標準モデルを用いた認識実験の結果を示すグラフである。

図 3 9 は、第 3 の実施の形態における第 2 近似部により作成された標準モデルによる認識実験の結果を示すグラフである。

図 4 0 は、本発明の第 9 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。

図 4 1 は、アプリ・仕様情報対応データベースのデータ例を示す図である。

図 4 2 は、PDA 1 0 0 1 の動作手順を示すフローチャートである。

図 4 3 は、参照モデル記憶部に格納されている参照モデルの一例を示す図である。

図 4 4 は、初期標準モデル作成部によるクラスタリングによる初期値の決定方法を示すフローチャートである。

5 図 4 5 は、図 4 4 におけるステップ S 1 0 0 4 の具体例を示す図である。

図 4 6 は、図 4 4 におけるステップ S 1 0 0 5 の具体例を示す図である。

10 図 4 7 は、図 4 4 におけるステップ S 1 0 0 6 の具体例を示す図である。

図 4 8 は、図 4 4 におけるステップ S 1 0 0 8 の具体例を示す図である。

図 4 9 は、本発明の第 1 0 の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

15 図 5 0 は、同サーバの動作手順を示すフローチャートである。

図 5 1 は、本発明に係る標準モデル作成装置を具体的に適用したシステム例を示す図である。

図 5 2 は、クラス I D ・ 初期標準モデル ・ 参照モデル対応表の例を示す図である。

20 図 5 3 は、図 5 2 のクラス I D ・ 初期標準モデル ・ 参照モデル対応表における参照モデル 8 A A ~ A Z の例を示す図である。

図 5 4 は、図 5 2 のクラス I D ・ 初期標準モデル ・ 参照モデル対応表における参照モデル 6 4 Z A ~ Z Z の例を示す図である。

25 図 5 5 は、図 5 2 のクラス I D ・ 初期標準モデル ・ 参照モデル対応表における初期標準モデル 8 A ~ 6 4 Z の例を示す図である。

図 5 6 は、クラス I D ・ 初期標準モデル ・ 参照モデル対応表の作成方

法を示すフローチャートである。

図 5 7 は、図 5 6 におけるステップ S 1 1 0 0 の具体例を示す図である。

5 図 5 8 は、図 5 6 におけるステップ S 1 1 0 2 の具体例を示す図である。

図 5 9 は、図 5 6 におけるステップ S 1 1 0 3 の具体例を示す図である。

図 6 0 は、図 5 6 におけるステップ S 1 1 0 4 の具体例を示す図である。

10 図 6 1 は、端末がサーバと通信することによってクラス I D ・ 初期標準モデル・参照モデル対応表を完成させる手順を示す図である。

図 6 2 は、クラス I D ・ 初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定方法を示すフローチャートである。

15 図 6 3 は、図 6 2 におけるステップ S 1 1 0 5 の具体例を示す図である。

図 6 4 は、第 3 近似部を用いて作成した標準モデルを用いた認識実験の結果を示すグラフである。

図 6 5 (a) ~ (j) は、音声認識の対象についての属性と標準モデルの構造 (ガウス分布の混合数) との関係例を示す図である。

20

発明を実施するための最良の形態

以下、本発明の実施の形態について図面を参照しながら詳しく説明する。なお、図中同一又は相当部分には同一符号を付し、その説明は繰り返さない。

25 (第 1 の実施の形態)

図 1 は、本発明の第 1 の実施の形態における標準モデル作成装置の全

体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ１０１に組み込まれた例が示されている。本実施の形態では特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する場合を例にして説明する。

サーバ１０１は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部１１１と、参照モデル準備部１０２と、参照モデル記憶部１０３と、標準モデル作成部１０４と、書き込み部１１２とを備える。

読み込み部１１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた子供用参照モデル、成人用参照モデル、高齢者用参照モデルを読み込む。参照モデル準備部１０２は、読み込まれた参照モデル１２１を参照モデル記憶部１０３へ送信する。参照モデル記憶部１０３は、３個の参照モデル１２１を記憶する。ここで、参照モデルとは、標準モデルを作成するに際して参照される予め作成されたモデル（ここでは、音声認識用のモデル、つまり、一定の属性を有する音声の特徴を示す確率モデル）である。

標準モデル作成部１０４は、参照モデル記憶部１０３が記憶した３個（ $N_g = 3$ ）の参照モデル１２１に対する確率又は尤度を最大化又は極大化するように標準モデル１２２を作成する処理部であり、標準モデルの構造（ガウス分布の混合数など）を決定する標準モデル構造決定部１０４ａと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部１０４ｂと、決定された初期標準モデルを記憶する統計量記憶部１０４ｃと、統計量記憶部１０４ｃに記憶された初期標準モデルに対して、第１近似部１０４ｅに

よる近似計算等を用いることにより、参照モデル記憶部 103 に記憶されている 3 個 ($N_g = 3$) の参照モデル 121 に対する確率又は尤度を最大化又は極大化するような統計量を算出する (最終的な標準モデルを生成する) 統計量推定部 104 d とからなる。なお、統計量とは、標準
5 モデルを特定するパラメータであり、ここでは、混合重み係数、平均値、分散値である。

書き込み部 112 は、標準モデル作成部 104 が作成した標準モデル 122 を CD-ROM などのストレージデバイスに書き込む。

次に、以上のように構成されたサーバ 101 の動作について説明する。

10 図 2 は、サーバ 101 の動作手順を示すフローチャートである。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する (ステップ S100)。つまり、読み込み部 111 は、CD-ROM などのストレージデバイスに書き込まれた子供用参照モデル、成人用参照モデル、高齢者用参照モデルを読み込み、参照モデル準備部 102 は、
15 読み込まれた参照モデル 121 を参照モデル記憶部 103 へ送信し、参照モデル記憶部 103 は、3 個の参照モデル 121 を記憶する。

参照モデル 121 は、音素ごとの HMM により構成される。参照モデル 121 の一例を図 3 に示す。ここでは、子供用参照モデル、成人用参照モデル、高齢者用参照モデルのイメージ図が示されている (なお、本
20 図では、高齢者用参照モデルのイメージ図は省略されている)。これら 3 個の参照モデルの全てが、状態数 3 個、各状態は混合分布数が 3 個の混合ガウス分布により HMM の出力分布が構成される。特徴量として 12 次元 ($J = 12$) のケプストラム係数が用いられる。

次に、標準モデル作成部 104 は、参照モデル記憶部 103 が記憶した 3 個の参照モデル 121 に対する確率又は尤度を最大化又は極大化す
25 るように標準モデル 122 を作成する (ステップ S101)。

最後に、書き込み部 1 1 2 は、標準モデル作成部 1 0 4 が作成した標準モデル 1 2 2 を C D - R O M などのストレージデバイスに書き込む（ステップ S 1 0 2）。C D - R O M などのストレージデバイスに書き込まれた標準モデルは、子供、成人、高齢者を考慮した音声認識用の標準
5 モデルとして利用される。

図 4 は、図 2 におけるステップ S 1 0 1（標準モデルの作成）の詳細な手順を示すフローチャートである。

まず、標準モデル構造決定部 1 0 4 a は、標準モデルの構造を決定する（ステップ S 1 0 2 a）。ここでは、標準モデルの構造として、音素ご
10 との H M M により構成され、3 状態であり、各状態における出力分布の混合数を 3 個（ $M_f = 3$ ）と決定する。

次に、初期標準モデル作成部 1 0 4 b は、標準モデルを計算するための統計量の初期値を決定する（ステップ S 1 0 2 b）。ここでは、参照モデル記憶部 1 0 3 に記憶された 3 つの参照モデルを、統計処理計算を用
15 いて 1 つのガウス分布に統合したものを統計量の初期値とし、その初期値を初期標準モデルとして統計量記憶部 1 0 4 c に記憶する。

具体的には、初期標準モデル作成部 1 0 4 b は、上記 3 つの状態 I （ $I = 1, 2, 3$ ）それぞれについて、以下の式 1 3 に示される出力分布を生成する。なお、式中の M_f （ガウス分布の混合数）は、ここでは、3 で
20 ある。

（式 1 3）

$$\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)$$

ここで、

（式 1 4）

$$f(x; \mu_{f(m)}, \sigma_{f(m)}^2) \quad (m = 1, 2, \dots, M_f)$$

は、ガウス分布を表し、

(式 15)

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

5 は、12次元 ($J = 12$) のLPCケプストラム係数を表し、

(式 16)

$$\omega_{f(m)} \quad (m = 1, 2, \dots, M_f)$$

は、各ガウス分布の混合重み係数を表し、

(式 17)

$$\mu_{f(m)} = (\mu_{f(m,1)}, \mu_{f(m,2)}, \dots, \mu_{f(m,J)}) \in R^J \quad (m = 1, 2, \dots, M_f)$$

10

は、各ガウス分布の平均値を表し、

(式 18)

$$\sigma_{f(m)}^2 = (\sigma_{f(m,1)}^2, \sigma_{f(m,2)}^2, \dots, \sigma_{f(m,J)}^2) \in R^J \quad (m = 1, 2, \dots, M_f)$$

は、各ガウス分布の分散値を表す。

15 そして、統計量推定部 104d は、参照モデル記憶部 103 に記憶された 3 つの参照モデル 121 を用いて、統計量記憶部 104c に記憶された標準モデルの統計量を推定する (ステップ S102c)。

具体的には、3 つ ($N_g = 3$) の参照モデル 121 の各状態 I ($I = 1, 2, 3$) における出力分布、即ち、以下の式 19 に示される出力分布に対する標準モデルの確率又は尤度 (以下の式 25 に示される尤度 $\log P$) を極大化もしくは最大化するような標準モデルの統計量 (上記式

20

16 に示される混合重み係数、上記式 17 に示される平均値、及び、上記式 18 に示される分散値) を推定する。

(式 19)

$$\sum_{l=1}^{L_{g(i)}} \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \quad (i=1,2,\dots,N_g)$$

5 ここで、

(式 20)

$$g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \quad (i=1,2,\dots,N_g, l=1,2,\dots,L_{(i)})$$

はガウス分布を表し、

(式 21)

$$L_{g(i)} \quad (i=1,2,\dots,N_g)$$

10

は各参照モデルの混合分布数 (ここでは、3) を表し、

(式 22)

$$\nu_{g(i,l)} \quad (l=1,2,\dots,L_{g(i)})$$

は各ガウス分布の混合重み係数を表し、

15 (式 23)

$$\mu_{g(i,l)} \quad (l=1,2,\dots,L_{g(i)})$$

は各ガウス分布の平均値を表し、

(式 24)

$$\sigma_{g(i,l)}^2 \quad (l=1,2,\dots,L_{g(i)})$$

は各ガウス分布の分散値を表す。

(式 2 5)

$$\log P = \sum_{l=1}^{N_g} \int_{-\infty}^{\infty} \log \left[\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2) \right] \left\{ \sum_{l=1}^{L_g(l)} \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx$$

そして、以下の式 2 6、式 2 7 及び式 2 8 に従って、それぞれ、標準
5 モデルの混合重み係数、平均値及び分散値を算出する。

(式 2 6)

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) \left\{ \sum_{l=1}^{L_g(l)} \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, k) \left\{ \sum_{l=1}^{L_g(l)} \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

($m = 1, 2, \dots, M_f$)

(式 2 7)

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) x_{(j)} \left\{ \sum_{l=1}^{L_g(l)} \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) \left\{ \sum_{l=1}^{L_g(l)} \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

($m = 1, 2, \dots, M_f, j = 1, 2, \dots, J$)

10 (式 2 8)

$$\sigma_{f(m,j)}^2 = \frac{\sum_{l=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) (x_{(j)} - \mu_{f(m,j)})^2 \left\{ \sum_{l=1}^{L_g(l)} v_{g(l,j)} g(x; \mu_{g(l,j)}, \sigma_{g(l,j)}^2) \right\} dx}{\sum_{l=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) \left\{ \sum_{l=1}^{L_g(l)} v_{g(l,j)} g(x; \mu_{g(l,j)}, \sigma_{g(l,j)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

このとき、統計量推定部 104d の第 1 近似部 104e により、以下の式 29 に示される近似式が用いられる。

(式 29)

$$\gamma(x, m) = \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x; \mu_{f(k)}, \sigma_{f(k)}^2)} \approx \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)}{u_{h(m)} h(x; \mu_{h(m)}, \sigma_{h(m)}^2)}$$

$$(m = 1, 2, \dots, M_f)$$

5

ここで、

(式 30)

$$u_{h(m)} h(x; \mu_{h(m)}, \sigma_{h(m)}^2) \quad (m = 1, 2, \dots, M_f)$$

は、

10 (式 31)

$$u_{h(m)} \quad (m = 1, 2, \dots, M_f)$$

(式 32)

$$\mu_{h(m)} = (\mu_{h(m,1)}, \mu_{h(m,2)}, \dots, \mu_{h(m,J)}) \in R^J$$

を平均値とし、

(式 3 3)

$$\sigma_{h(m)}^2 = (\sigma_{h(m,1)}^2, \sigma_{h(m,2)}^2, \dots, \sigma_{h(m,J)}^2) \in R^J$$

を分散値とする単一のガウス分布を表す。

- 5 また、第 1 近似部 104 e は、上記式 30 に示された単一ガウス分布の重み (式 3 1) 平均値 (式 3 2) 及び分散値 (式 3 3) を、それぞれ、以下の式 3 4、式 3 5 及び式 3 6 に示された式に従って算出する。

(式 3 4)

$$u_{h(m)} = \sum_{p=1}^{M_f} \omega_{f(m,p)} = \sum_{p=1}^{M_f} \omega_{f(p)} = 1.0 \quad (m=1,2,\dots,M_f)$$

10 (式 3 5)

$$\mu_{h(m,j)} = \frac{\sum_{p=1}^{M_f} \omega_{f(m,p)} \mu_{f(m,p,j)}}{\sum_{p=1}^{M_f} \omega_{f(m,p)}} = \frac{\sum_{p=1}^{M_f} \omega_{f(p)} \mu_{f(p,j)}}{\sum_{p=1}^{M_f} \omega_{f(p)}}$$

$$(m=1,2,\dots,M_f, j=1,2,\dots,J)$$

(式 3 6)

$$\sigma_{h(m,j)}^2 = \frac{\sum_{p=1}^{M_f} \omega_{f(m,p)} (\sigma_{f(m,p,j)}^2 + \mu_{f(m,p,j)}^2)}{\sum_{p=1}^{M_f} \omega_{f(m,p)}} - \mu_{h(m,j)}^2$$

$$= \frac{\sum_{p=1}^{M_f} \omega_{f(p)} (\sigma_{f(p,j)}^2 + \mu_{f(p,j)}^2)}{\sum_{p=1}^{M_f} \omega_{f(p)}} - \mu_{h(m,j)}^2$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

図 5 は、第 1 近似部 104 e による近似計算を説明する図である。第 1 近似部 104 e は、本図に示されるように、上記式 29 に示された近似式における単一ガウス分布（式 30）を、標準モデルを構成する全ての混合ガウス分布を用いて決定している。

以上の第 1 近似部 104 e による近似式を考慮してまとめると、統計量推定部 104 d での計算式は次の通りになる。つまり、統計量推定部 104 d は、以下の式 37、式 38 及び式 39 に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、統計量記憶部 104 c に記憶する。そして、このような統計量の推定と統計量記憶部 104 c への記憶を R（ ≥ 1 ）回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル 122 の統計量として出力する。

（式 37）

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \prod_{j=1}^J \sum_{l=1}^{L_g(i)} A_{(m,i,l,j)}}{\sum_{i=1}^{N_g} \sum_{k=1}^{M_f} \prod_{j=1}^J \sum_{l=1}^{L_g(i)} A_{(k,i,l,j)}} \quad (m=1,2,\dots,M_f)$$

$$A_{(m,i,l,j)} = \frac{\omega_{f(m)} \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2}{\sqrt{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2}} \\ \times \exp \left\{ \frac{1}{2} \left(\frac{\left(\frac{\sigma_{f(m,j)} \sigma_{g(i,l,j)} \mu_{h(m,j)} - \frac{\sigma_{f(m,j)} \sigma_{h(m,j)} \mu_{g(i,l,j)} - \frac{\sigma_{g(i,l,j)} \sigma_{h(m,j)} \mu_{f(m,j)}}{\sigma_{h(m,j)}} \right)^2}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2} + \frac{\mu_{h(m,j)}}{\sigma_{h(m,j)}^2} - \frac{\mu_{g(i,l,j)}}{\sigma_{g(i,l,j)}^2} - \frac{\mu_{f(m,j)}}{\sigma_{f(m,j)}^2} \right) \right\}$$

(式 38)

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} B_{(m,i,l,j)}}{\sum_{i=1}^{N_g} \prod_{j=1}^J \sum_{l=1}^{L_g(i)} A_{(m,i,l,j)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

$$B_{(m,i,l,j)} = \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(i,l,j)} + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2} \times A_{(m,i,l,j)}$$

(式 39)

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} C_{(m,i,l,j)}}{\sum_{i=1}^{N_g} \prod_{j=1}^J \sum_{l=1}^{L_g(i)} A_{(m,i,l,j)}} \quad (m=1,2,\dots,M_f)$$

$$C_{(m,i,l,j)} = \left\{ \frac{\sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2} + \left(\mu_{f(m,j)} - \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(i,l,j)} + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2} \right)^2 \right\} \times A_{(m,i,l,j)}$$

なお、状態遷移確率については、HMMの対応する状態遷移確率を参照モデル 1 2 1 に対して全て加えあわせた全体が 1 になるように正規化したものを用いる。

次に、本実施の形態をパーソナルコンピュータによる音声認識に適用した具体例を説明する。ここでは、サーバ 1 0 1 としてパソコン(PC)、読み込み部 1 1 1 としてCD-ROMドライブ装置を用いるものとし、標準モデルの具体的な使い方を中心に説明する。

まず、利用者は、PC(サーバ 1 0 1)のCD-ROMドライブ装置(読み込み部 1 1 1)に、参照モデルとしての複数の音響モデルが格納された 1 枚のCD-ROMを装着する。そのCD-ROMには、例えば、「幼児」、「子供：男」、「子供：女」、「大人：男」、「大人：女」、「高齢者：男」、「高齢者：女」の各音響モデルが記憶されている。

次に、利用者は、図 6 (a) 及び (b) に示される画面表示例のように、PC(サーバ 1 0 1)に接続されたディスプレイを用いて、家族構成(音声認識を利用する人)にあった音響モデルを選択する。図 6 には、CD-ROMに記憶されている音響モデルが「CD-ROM」と書かれた枠内に表示され、それらの音響モデルの中から選択された音響モデルが「利用者」と書かれた枠内にコピーされる様子が示されている。ここでは、利用者の家族構成が、10歳の男の子と、50歳のお父さんと、40歳のお母さんの3人であるとし、利用者(お父さん)によって、「子供：男」、「大人：男」、「大人：女」の3個のモデルが「利用者」と書かれた枠内にドラッグして移動されている。このような操作によって、参照モデル準備部 1 0 2 による参照モデルの準備が行われる。つまり、3個の参照モデルが読み込み部 1 1 1 で読み出され、参照モデル準備部 1 0 2 を介して、参照モデル記憶部 1 0 3 に格納される。

続いて、利用者は、図 7 (a) に示される画面表示例のように、作成

する標準モデルの構造（混合分布数）を指定する。図 7（a）では、「混合分布数」として「3 個」、「10 個」、「20 個」が表示され、利用者は、これらの個数の中から希望するものを選択する。この操作によって、標準モデル構造決定部 104 a により、これから作成する標準モデルの構造が決定される。

なお、混合分布数の決定については、このような直接的な指定に限られず、例えば、図 7（b）に示される画面表示例のように、利用者が選択した仕様情報に基づいて混合分布数を決定してもよい。図 7（b）では、標準モデルを使用して音声認識を実行させる対象機器として、3 種類の「利用機器」、つまり、「テレビ用」、「カーナビ用」、「携帯電話用」の中から利用機器を選択する様子が示されている。このとき、予め記憶された対応表に従って、例えば、「テレビ用」が選択された場合には混合分布数を 3 個と決定し、「カーナビ用」が選択された場合には混合分布数を 20 個と決定し、「携帯電話用」が選択された場合には混合分布数を 10 個と決定してもよい。

その他、混合分布数の決定については、認識速度や精度、つまり、「素早く認識」、「通常」、「高精度に認識」の中から選択することで、それぞれの選択項目に対応した値（「素早く認識」＝3 個、「通常」＝10 個、「高精度に認識」＝20 個）を混合分布数として決定してもよい。

このような入力操作が終了すると、初期標準モデル作成部 104 b によって初期標準モデルが作成された後に、統計量推定部 104 d による繰り返し計算（学習）が行われ、標準モデルが作成される。このとき、図 8 の画面表示例に示されるように、標準モデル構造決定部 104 a によって、学習の進捗状況が表示される。利用者は、学習の進捗状況、学習終了時期などを知ることができ、標準モデルが完成されるまで安心して待つことができる。なお、進捗状況の表示として、例えば、図 8（a）

に示されるような学習度合いのバー表示、図 8 (b) に示されるような学習回数の表示、その他、尤度基準の表示等がある。また、未学習時は一般的な顔画像を表示し、学習の完了に近づくにしがって利用者の顔画像に変更していくような進捗表示であってもよい。同様に、未学習時
5 には赤ちゃんを表示し、学習の完了に近づくにしがって仙人を表示するような進捗表示であってもよい。

このようにして標準モデルの作成が完了すると、作成された標準モデルは、標準モデル作成部 104 によってメモリカード（書き込み部 112）に記録される。利用者は、そのメモリカードを PC（サーバ 101）の書き込み部 112 から抜き出し、利用機器、例えば、テレビのメモリカード用スロットに挿入する。これによって、作成された標準モデルが PC（サーバ 101）から利用機器（テレビ）に移動される。テレビは、装着されたメモリカードに記録された標準モデルを用いて、利用者（ここでは、テレビを利用する家族）を対象とした音声認識を行う。た
10 とえば、テレビに付属したマイクに入力された音声を認識することによって、テレビ操作のコマンドを判別し、そのコマンド（例えば、チャンネルの切り替え、EPG などによる番組検索）を実行する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた、音声によるテレビ操作が実現される。

20 以上説明したように、本発明の第 1 の実施の形態によれば、予め準備された参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、学習のためのデータや教師データを必要とすることなく簡易に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した
25 精度の高い標準モデルが作成される。

なお、標準モデル 122 は、音素ごとに HMM を構成するに限らず、

文脈依存のHMMで構成してもよい。

また、標準モデル作成部104は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル122を構成するHMMは、音素ごとに異なる状態
5 数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル121は、子供用参照モデル、成人用参照モデル、高齢者用参照モデルにおいて、異なる状態数により構成してもよいし、異なる混合数の混合ガウス分布により構成してもよい。

10 また、標準モデル122を用いて、サーバ101において音声認識を行ってもよい。

また、参照モデル121をCD-ROM、DVD-RAMなどのストレージデバイスから読み込む代わりに、サーバ101において音声データから参照モデル121を作成してもよい。

15 また、参照モデル準備部102は、必要に応じてCD-ROM、DVD-RAMなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部103に追加・更新してもよい。つまり、参照モデル準備部102は、新たな参照モデルを参照モデル記憶部103に格納するだけでなく、同一の認識対象についての参照モデルが参照モデル
20 記憶部103に格納されている場合には、その参照モデルと置き換えることによって参照モデルを更新したり、参照モデル記憶部103に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部102は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部103に追加・更新してもよい。

25 また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

(第2の実施の形態)

図9は、本発明の第2の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がセットトップボックス201（以下、STBと呼ぶ）に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（話者適応モデル）を作成する場合を例にして説明する。具体的には、STBによる音声認識機能により、テレビのEPG検索や番組切替、録画予約などを行う場合を例にして説明する。

STB201は、ユーザの発話を認識してTV番組の自動切替等を行うデジタル放送用受信機であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、マイク211と、音声データ蓄積部212と、参照モデル準備部202と、参照モデル記憶部203と、利用情報作成部204と、参照モデル選択部205と、標準モデル作成部206と、音声認識部213とを備える。

マイク211に収集された音声データは、音声データ蓄積部212に蓄積される。参照モデル準備部202は、音声データ蓄積部212が蓄積した音声データを用いて話者ごとに参照モデル221を作成し、参照モデル記憶部203に記憶する。

利用情報作成部204は、利用情報224である利用者の音声をマイク211により収集する。ここで、利用情報とは、認識（狭義での認識、識別、認証など）の対象（人・物）に関する情報であり、ここでは、音声認識の対象となる利用者の音声である。参照モデル選択部205は、利用情報作成部204が作成した利用情報224に基づいて、参照モデル

ル記憶部 203 が記憶している参照モデル 221 の中から、利用情報 224 が示す利用者の音声に音響的に近い参照モデル 223 を選択する。

標準モデル作成部 206 は、参照モデル選択部 205 が選択した話者の参照モデル 223 に対する確率又は尤度を最大化又は極大化するように標準モデル 222 を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部 206a と、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部 206b と、決定された初期標準モデルを記憶する統計量記憶部 206c と、統計量記憶部 206c に記憶された初期標準モデルに対して、一般近似部 206e による近似計算等を用いることにより、参照モデル選択部 205 が選択した参照モデル 223 に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部 206d とからなる。

15 音声認識部 213 は、標準モデル作成部 206 によって作成された標準モデル 222 を用いて利用者の音声を認識する。

次に、以上のように構成された STB 201 の動作について説明する。

図 10 は、STB 201 の動作手順を示すフローチャートである。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップ S200）。つまり、マイク 211 により A さんから Z さんの音声データを収集して音声データ蓄積部 212 に蓄積する。たとえば、屋内に設置された複数のマイク、テレビのリモコンに内蔵されたマイク、電話機などが、STB 201 の音声データ蓄積部 212 と接続されており、マイクや電話機から入力された音声データを音声データ蓄積部 212 に蓄積する。たとえば、お兄ちゃん、妹、お父さん、お母さん、おじいちゃん、近所のひと、友達の音声蓄積される。

参照モデル準備部 202 は、音声データ蓄積部 212 が蓄積した音声データを用いて話者ごとに参照モデル 221 をバウム・ウェルチの再推定の方法により作成する。この処理は、標準モデルの作成が要求される以前に行われる。

- 5 参照モデル記憶部 203 は、参照モデル準備部 202 が作成した参照モデル 221 を記憶する。参照モデル 221 は、音素ごとの HMM により構成される。参照モデル 221 の一例を図 11 に示す。ここでは、A さんから Z さんの全ての参照モデルが、状態数 3 個、各状態は混合分布数が 5 個の混合ガウス分布により HMM の出力分布が構成される。特徴
- 10 量として 25 次元 ($J=25$) のメルケプストラム係数が用いられる。

ここで、標準モデルの作成が要求される。たとえば、利用者が「利用者の確認」のボタンを押すことによって、標準モデルの作成が要求される。「利用者確認」のボタンについては、テレビ画面に表示させて選択する方法や、テレビのリモコンに「利用者の確認」スイッチをつけて選択

15 する方法が考えられる。ボタンを押すタイミングとしては、テレビを起動したタイミング、音声認識を用いてコマンド操作を行っているときに利用者にふさわしい標準モデルがほしいと感じたタイミングなどが考えられる。

- 次に、利用情報作成部 204 は、利用情報 224 である利用者の音声
- 20 をマイク 211 により収集する (ステップ S201)。たとえば、標準モデルの作成が要求されると、画面上で「名前を入力してください」と表示される。利用者は、テレビのリモコンに内蔵されたマイクにより名前 (利用者の音声) を入力する。この利用者の音声を利用情報である。なお、入力する音声は名前に限定されない。例えば「適応と発声してください」と表示して、利用者は「適応」と発声してもよい。
- 25

参照モデル選択部 205 は、その利用者の音声に音響的に近い参照モ

デル 2 2 3 を、参照モデル記憶部 2 0 3 が記憶している参照モデル 2 2 1 の中から選択する（ステップ S 2 0 2）。具体的には、利用者の音声を A さんから Z さんの参照モデルに入力して発声単語に対する尤度が大きい 1 0 人（ $N_g = 10$ ）の話者の参照モデルを選択する。

- 5 そして、標準モデル作成部 2 0 6 は、参照モデル選択部 2 0 5 が選択した 1 0 個の参照モデル 2 2 3 に対する確率又は尤度を最大化又は極大化するように標準モデル 2 2 2 を作成する（ステップ S 2 0 3）。このとき、第 1 の実施の形態のように、学習の進捗状況を表示してもよい。そうすることで、利用者は学習の進捗状況、学習終了時期などが判断でき、
- 10 安心して標準モデルを作成することができる。また、学習の進捗状況为非表示にする進捗状況非表示部を設けてもよい。この機能により、画面を有効に使うことができる。また、慣れた人に対して非表示にすることで、うっとうしく感じるものが回避される。

- 最後に、音声認識部 2 1 3 は、マイク 2 1 1 から介して送られてくる
- 15 利用者の音声を入力とし、標準モデル作成部 2 0 6 で作成された標準モデル 2 2 2 を用いて音声認識を行う（S 2 0 4）。たとえば、利用者が発話した音声を音響解析等を行うことで 2 5 次元のメルケプストラム係数を算出し、音素ごとの標準モデル 2 2 2 に入力することで、高い尤度を有する音素の連なりを特定する。そして、その音素の連なりと予め受信
- 20 している電子番組データ中の番組名とを比較し、一定以上の尤度が検出された場合に、その番組に切り替えるという自動番組切替の制御を行う。

- 次に、図 1 0 におけるステップ S 2 0 3（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図 4 に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が
- 25 異なる。

まず、標準モデル構造決定部 2 0 6 a は、標準モデルの構造を決定す

る（図４のステップＳ１０２ａ）。ここでは、標準モデルの構造として、音素ごとのＨＭＭにより構成され、３状態であり、各状態における出力分布の混合分布数が１６個（ $Mf=16$ ）と決定する。

次に、初期標準モデル作成部２０６ｂは、標準モデルを計算するための統計量の初期値を決定する（図４のステップＳ１０２ｂ）。ここでは、参照モデル選択部２０５が選択した１０個の参照モデル２２３を、統計処理計算を用いて１つのガウス分布に統合したものを統計量の初期値とし、その初期値を初期標準モデルとして統計量記憶部２０６ｃに記憶する。ここでは、話者ごとに学習した混合分布数が５の参照モデルを用いて精度の高い混合分布数が１６（１６混合）の標準モデル（話者適応モデル）を作成する。

具体的には、初期標準モデル作成部２０６ｂは、上記３つの状態１（ $i=1, 2, 3$ ）それぞれについて、上記式１３に示される出力分布を生成する。

ただし、本実施の形態では、上記式１３に示された出力分布における（式４０）

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

は、２５次元（ $J=25$ ）のメルケプストラム係数を表す。

そして、統計量推定部２０６ｄは、参照モデル選択部２０５が選択した１０個の参照モデル２２３を用いて、統計量記憶部２０６ｃに記憶された標準モデルの統計量を推定する（図４のステップＳ１０２ｃ）。

つまり、１０個（ $Ng=10$ ）の参照モデル２２３の各状態１（ $i=1, 2, 3$ ）における出力分布、即ち、上記式１９に示される出力分布に対する標準モデルの確率（ここでは、上記式２５に示される尤度 $\log P$ ）を極大化もしくは最大化するような標準モデルの統計量（上記式１

6 に示される混合重み係数、上記式 17 に示される平均値、及び、上記式 18 に示される分散値) を推定する。

ただし、本実施の形態では、上記式 19 に示された出力分布における (式 41)

$$L_{g(i)} \quad (i=1,2,\dots,N_g)$$

5

は、5 (各参照モデルの混合分布数) である。

具体的には、上記式 26、式 27 及び式 28 に従って、それぞれ、標準モデルの混合重み係数、平均値及び分散値を算出する。

このとき、統計量推定部 206d の一般近似部 206e により、上記式 29 に示される近似式が用いられる。

ここで、一般近似部 206e は、第 1 の実施の形態と異なり、上記式 29 の近似式の分母に示された出力分布 (式 42)

$$\omega_{f(k)} f(x; \mu_{f(k)}, \sigma_{f(k)}^2) \quad (k=1,2,\dots,M_f)$$

15 の中から、上記式 29 の近似式の分子に示された出力分布 (式 43)

$$\omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)$$

に距離的に近い 3 個 ($P_{h(m)}=3$) の出力分布 (式 44)

$$\omega_{f(m,p)} f(x; \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m=1,2,\dots,M_f, p=1,2,\dots,P_{h(m)})$$

20

を選択し、選択した 3 個の出力分布を用いて、上記式 30 に示された単一ガウス分布の重み (式 31)、平均値 (式 32) 及び分散値 (式 33)

を、それぞれ、以下の式 4 5、式 4 6 及び式 4 7 に示された式に従って算出する。

(式 4 5)

$$u_{h(m)} = \sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \quad (m=1,2,\dots,M_f)$$

5 (式 4 6)

$$\mu_{h(m,j)} = \frac{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \mu_{f(m,p,j)}}{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

(式 4 7)

$$\sigma_{h(m,j)}^2 = \frac{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} (\sigma_{f(m,p,j)}^2 + \mu_{f(m,p,j)}^2)}{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)}} - \mu_{h(m,j)}^2$$

$$(m=1,2,\dots,M_f, j=1,2,\dots,J)$$

- 10 図 1 2 は、一般近似部 2 0 6 e による近似計算を説明する図である。
 一般近似部 2 0 6 e は、本図に示されるように、上記式 2 9 に示された
 近似式における単一ガウス分布 (式 3 0) を、標準モデルを構成する Mf
 個の混合ガウス分布の中から、計算対象となる混合ガウス分布に近い一
 部 (P_{h(m)} 個) の混合ガウス分布だけを用いて決定している。したがっ
 15 て、全部 (Mf 個) の混合ガウス分布を用いる第 1 の実施の形態と比較
 し、近似計算における計算量が削減される。

以上の一般近似部 2 0 6 e による近似式を考慮してまとめると、統計
 量推定部 2 0 6 d での計算式は次の通りになる。つまり、統計量推定部

206dは、以下の式48、式49及び式50に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、統計量記憶部206cに記憶する。そして、このような統計量の推定と統計量記憶部206cへの記憶をR（ ≥ 1 ）回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル222の統計量として出力する。なお、繰り返し計算においては、その回数に対応させて、上記近似計算における出力分布の選択個数Ph(m)を小さくし、最終的にPh(m)=1とする計算を行う。

(式48)

$$\omega_{f(m)} = \frac{\sum_{l=1}^{N_g} \sum_{i=1}^{L_{g(l)}} \alpha_{(m,l,i)}}{\sum_{k=1}^{M_f} \omega_{f(k)} \left(\sum_{l=1}^{N_g} \sum_{i=1}^{L_{g(l)}} \alpha_{(k,l,i)} \right)} \quad (m=1,2,\dots,M_f)$$

$$\alpha_{(m,l,i)} = v_{g(l,i)} \prod_{j=1}^J D_{(m,l,i,j)}$$

$$D_{(m,l,i,j)} = \frac{\sigma_{H(m,j)}^2}{\sqrt{\sigma_{f(m,j)}^2 \sigma_{H(m,j)}^2 + \sigma_{g(l,j)}^2 \sigma_{H(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2}} \times \exp \left\{ \frac{1}{2} \left[\left(\frac{\sigma_{f(m,j)} \sigma_{g(l,j)}}{\sigma_{H(m,j)}} \mu_{H(m,j)} - \frac{\sigma_{f(m,j)} \sigma_{H(m,j)}}{\sigma_{g(l,j)}} \mu_{g(l,j)} - \frac{\sigma_{g(l,j)} \sigma_{H(m,j)}}{\sigma_{f(m,j)}} \mu_{f(m,j)} \right)^2 \right. \right. \\ \left. \left. + \frac{\mu_{H(m,j)}}{\sigma_{H(m,j)}^2} \frac{\mu_{g(l,j)}}{\sigma_{g(l,j)}^2} \frac{\mu_{f(m,j)}}{\sigma_{f(m,j)}^2} \right] \right\}$$

10 (式49)

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(l)} \beta_{(m,l,i,j)} \alpha_{(m,l,i)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(l)} \alpha_{(m,l,i)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

$$\beta_{(m,l,i,j)} = \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(l,m,j)} + \sigma_{g(l,i,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(l,i,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,i,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,i,j)}^2}$$

(式 50)

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(l)} \gamma_{(m,l,i,j)} \alpha_{(m,l,i)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(l)} \alpha_{(m,l,i)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

$$\gamma_{(m,l,i,j)} = \left\{ \frac{\sigma_{f(m,j)}^2 \sigma_{g(l,i,j)}^2 \sigma_{h(m,j)}^2}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,i,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,i,j)}^2} + \left(\mu_{f(m,j)} - \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(l,m,j)} + \sigma_{g(l,i,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(l,i,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,i,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,i,j)}^2} \right)^2 \right\}$$

5 なお、状態遷移確率については、HMMの対応する状態遷移確率を参照モデル 223 に対して全て加えあわせた全体が 1 になるように正規化したものを用いる。

以上説明したように、本発明の第 2 の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成
10 されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

なお、標準モデルを作成するタイミングとしては、本実施の形態のよ

うな利用者による明示的な指示だけに限られず、他のタイミングで標準モデルを作成してもよい。たとえば、STB201にさらに、利用者が変更されたかどうかを自動的に判断する利用者変更判断部を設ける。その利用者変更判断部は、テレビのリモコンに入力された認識用の音声を用いて、利用者が変更されたか否か、つまり、現在の利用者が直前まで認識していた利用者と同一人物であるか否かを判断する。利用者が変更されたと判断した場合に、その音声を利用情報として標準モデルを作成する。これにより、利用者が意識することなく、利用者にふさわしい標準モデルを用いた音声認識が行われる。

- 5 なお、標準モデル222は、音素ごとにHMMを構成するに限らず、文脈依存のHMMで構成してもよい。

また、標準モデル作成部206は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

- 15 また、標準モデル222を構成するHMMは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル221は、話者ごとHMMにおいて、異なる状態数により構成してもよいし、異なる混合数の混合ガウス分布により構成してもよい。

- 20 また、参照モデル221は、話者ごとHMMに限らず、話者・雑音・声の調子ごとに作成してもよい。

また、標準モデル222をCD-ROM、ハードディスク、DVD-RAMなどのストレージデバイスに記録してもよい。

- 25 また、参照モデル221を作成する代わりに、CD-ROM、DVD-RAMなどのストレージデバイスから読み込んでもよい。

また、参照モデル選択部205は、利用情報224に基づいて利用者

ごとに選択する参照モデルの数を変えてもよい。

また、参照モデル準備部 202 は、必要に応じて新たな参照モデルを作成して参照モデル記憶部 203 に追加・更新してもよいし、参照モデル記憶部 203 に格納されている不要な参照モデルを削除してもよい。

- 6 また、参照モデル準備部 202 は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部 203 に追加・更新してもよい。

また、上記近似計算において選択する出力分布の個数 $P_h(m)$ は、対象とする事象や標準モデルの出力分布によって異なってもよいし、分布間距離に基づいて決定してもよい。

- 10 また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などの HMM の構造や、状態数などを決定してもよい。

- 15 また、混合分布数については、本実施の形態における STB を出荷するときに、所定の値に設定しておいてもよいし、ネットワーク連携を考慮した機器の CPU パワーなどの仕様、起動するアプリケーションの仕様などに基づいて混合分布数を決定してもよい。

(第 3 の実施の形態)

- 20 図 13 は、本発明の第 3 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置が PDA (Personal Digital Assistant) 301 に組み込まれた例が示されている。本実施の形態では雑音識別用の標準モデル (雑音モデル) を作成する場合を例にして説明する。

- 25 PDA 301 は、携帯情報端末であり、事象の出力確率によって定義される雑音識別用の標準モデルを作成する標準モデル作成装置として、

読み込み部 311 と、参照モデル準備部 302 と、参照モデル記憶部 303 と、利用情報作成部 304 と、参照モデル選択部 305 と、標準モデル作成部 306 と、仕様情報作成部 307 と、マイク 312 と、雑音識別部 313 とを備える。

- 5 読み込み部 311 は、CD-ROM などのストレージデバイスに書き込まれた乗用車 A の参照モデル、乗用車 B の参照モデル、バス A の参照モデル、小雨の参照モデル、大雨の参照モデルなどの雑音の参照モデルを読み込む。参照モデル準備部 302 は、読み込まれた参照モデル 321 を参照モデル記憶部 303 へ送信する。参照モデル記憶部 303 は、
- 10 参照モデル 321 を記憶する。

利用情報作成部 304 は、利用情報 324 である雑音の種類を PDA 301 の画面とキーを利用して作成する。参照モデル選択部 305 は、利用情報 324 である雑音の種類に音響的に近い参照モデルを、参照モデル記憶部 303 が記憶している参照モデル 321の中から選択する。

- 15 仕様情報作成部 307 は、PDA 301 の仕様に基づき仕様情報 325 を作成する。ここで、仕様情報とは、作成する標準モデルの仕様に関する情報であり、ここでは、PDA 301 が備える CPU の処理能力に関する情報である。

- 標準モデル作成部 306 は、仕様情報作成部 307 で作成された仕様
- 20 情報 325 に基づいて、参照モデル選択部 305 が選択した雑音の参照モデル 323 に対する確率又は尤度を最大化又は極大化するように標準モデル 322 を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部 306a と、標準モデルを計算するための統計量の初期値を決定することで初期標準モデル
- 25 ルを作成する初期標準モデル作成部 306b と、決定された初期標準モ

された初期標準モデルに対して、第2近似部306eによる近似計算等を用いることにより、参照モデル選択部305が選択した参照モデル323に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部306dとからなる。

雑音識別部313は、標準モデル作成部306で作成された標準モデル322を用いて、マイク312から入力された雑音の種類を識別する。

次に、以上のように構成されたPDA301の動作について説明する。

図14は、PDA301の動作手順を示すフローチャートである。

10 まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップS300）。つまり、読み込み部311は、ストレージデバイスに書き込まれた雑音の参照モデルを読み込み、参照モデル準備部302は、読み込まれた参照モデル321を参照モデル記憶部303へ送信し、参照モデル記憶部303は、参照モデル321を記憶する。

15 参照モデル321は、GMMより構成される。参照モデル321の一例を図15に示す。ここでは、各雑音モデルは混合分布数が3個のGMMにより構成される。特徴量として5次元（ $J=5$ ）のLPCケプストラム係数が用いられる。

次に、利用情報作成部304は、識別したい雑音の種類である利用情報324を作成する（ステップS301）。図16にPDA301の選択画面の一例を示す。ここでは、乗用車の雑音が選択される。参照モデル選択部305は、選択された利用情報324である乗用車の雑音に音響的に近い参照モデルである乗用車Aの参照モデルと乗用車Bの参照モデルを、参照モデル記憶部303が記憶している参照モデル321の中から
25 ら選択する（ステップS302）。

仕様情報 325 を作成する (ステップ S303)。ここでは、PDA301
の CPU の仕様に基づき CPU パワーが小さいという仕様情報 325 を
作成する。標準モデル作成部 306 は、作成された仕様情報 325 に基
づいて、参照モデル選択部 305 が選択した参照モデル 323 に対する
5 確率又は尤度を最大化又は極大化するように標準モデル 322 を作成す
る (ステップ S304)。

最後に、雑音識別部 313 は、利用者によってマイク 312 から入力
された雑音に対して、標準モデル 322 を用いて、雑音の識別を行う (ス
テップ S305)。

10 次に、図 14 におけるステップ S304 (標準モデルの作成) の詳細
な手順を説明する。手順の流れは、図 4 に示されたフローチャートと同
様である。ただし、採用する標準モデルの構造や具体的な近似計算等が
異なる。

まず、標準モデル構造決定部 306a は、標準モデルの構造を決定す
15 る (図 4 のステップ S102a)。ここでは、標準モデルの構造として、
仕様情報 325 である CPU パワーが小さいという情報に基づいて 1 混
合 ($M_f = 1$) の GMM により標準モデル 322 を構成すると決定する。

次に、初期標準モデル作成部 306b は、標準モデルを計算するため
の統計量の初期値を決定する (図 4 のステップ S102b)。ここでは、
20 選択された参照モデル 323 である乗用車 A の 3 混合の参照モデルを、
統計処理計算を用いて 1 つのガウス分布に統合したものを統計量の初期
値として統計量記憶部 306c に記憶する。

具体的には、初期標準モデル作成部 306b は、上記式 13 に示され
る出力分布を生成する。

25 ただし、本実施の形態では、上記式 13 に示された出力分布における
(式 51)

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

は、5次元（ $J=5$ ）のLPCケプストラム係数を表す。

そして、統計量推定部306dは、参照モデル選択部305が選択した2個の参照モデル323を用いて、統計量記憶部306cに記憶された標準モデルの統計量を推定する（図4のステップS102c）。

つまり、2個（ $N_g=2$ ）の参照モデル323における出力分布、即ち、上記式19に示される出力分布に対する標準モデルの確率（ここでは、上記式25に示される尤度 $\log P$ ）を極大化もしくは最大化するような標準モデルの統計量（上記式16に示される混合重み係数、上記式17に示される平均値、及び、上記式18に示される分散値）を推定する。

ただし、本実施の形態では、上記式19に示された出力分布における（式52）

$$L_{g(i)} \quad (i=1,2,\dots,N_g)$$

は、3（各参照モデルの混合分布数）である。

具体的には、上記式26、式27及び式28に従って、それぞれ、標準モデルの混合重み係数、平均値及び分散値を算出する。

このとき、統計量推定部306dの第2近似部306eは、標準モデルの各ガウス分布はお互いに影響を与えないと仮定して、以下の近似式を用いる。

（式53）

$$\gamma(x, m) \approx \frac{\omega_{f(m)} f(x, \mu_{f(m)}, \sigma_{f(m)}^2)}{u_{h(m)} h(x, \mu_{h(m)}, \sigma_{h(m)}^2)} \approx 1.0$$

$$(m = 1, 2, \dots, M_f)$$

また、標準モデルのガウス分布

(式 5 4)

$$\omega_{f(m,p)} f(x, \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m = 1, 2, \dots, M_f, p = 1, 2, \dots, P_{h(m)})$$

5 の近傍の

(式 5 5)

x

とは、前記式 5 4 が示す出力分布との平均値のユークリッド距離、マハラノビス距離、カルバック・ライブラー (KL) 距離などの分布間距

10 離が近い $Q_{g(m,i)}$ 個の参照モデル 3 2 3 のガウス分布

(式 5 6)

$$g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \quad (i = 1, 2, \dots, N_g, l = 1, 2, \dots, L_{(i)})$$

が存在する空間であって、

(式 5 7)

$$\omega_{f(m,p)} f(x, \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m = 1, 2, \dots, M_f, p = 1, 2, \dots, P_{h(m)})$$

15

との分布間距離が近い $Q_{g(m,i)}$ 個 ($1 \leq Q_{g(m,i)} \leq L_{g(i)}$) の前記参照ベクトルの出力分布とは、前記参照モデルの出力分布

(式 5 8)

$$v_{g(i,l)} g(x; \mu_{g(l)}, \sigma_{g(l)}^2) \quad (i = 1, 2, \dots, N_g, l = 1, 2, \dots, L_{g(i)})$$

のうち分布間距離が 1 番近い（近傍指示パラメータ $G = 1$ ）前記標準モデルの出力分布が前記式 57 である前記参照ベクトルの出力分布であると近似する。

5 図 17 は、この統計量推定部 306d による統計量の推定手順を示す概念図である。各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布を用いて統計量の推定を行うことが示されている。

10 図 18 は、第 2 近似部 306e による近似計算を説明する図である。第 2 近似部 306e は、本図に示されるように、各参照モデルの各ガウス分布に対して、距離が最も近い標準モデルのガウス分布 m を決定することで、上記式 53 に示された近似式を用いている。

15 以上の第 2 近似部 306e による近似式を考慮してまとめると、統計量推定部 306d での計算式は次の通りになる。つまり、統計量推定部 306d は、以下の式 59、式 60 及び式 61 に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、それらのパラメータによって特定される標準モデルを最終的な標準モデル 322 として生成する。

（式 59）

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)}}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)}}$$

$$(m = 1, 2, \dots, M_f)$$

(ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布に関する和を意味する。)

5 (式 60)

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,i)}} \nu_{g(i,l)} \mu_{g(i,l,j)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,i)}} \nu_{g(i,l)}}$$

$$(m=1,2,\dots,M_f, j=1,2,\dots,J)$$

(ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布に関する和を意味する。)

10

(式 61)

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,i)}} \nu_{g(i,l)} (\sigma_{g(i,l)}^2 + \mu_{g(i,l,j)}^2)}{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,i)}} \nu_{g(i,l)}} - \mu_{f(m,j)}^2$$

$$(m=1,2,\dots,M_f, j=1,2,\dots,J)$$

(ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布に関する和を

15

ただし、

(式 6 2)

$$\sum_{i=1}^{N_g} \mathcal{Q}_{g(m,i)} = 0 \quad (m = 1, 2, \dots, M_f)$$

の場合において、

5 (第 1 の方法) 混合重み係数、平均値、分散値を更新しない。

(第 2 の方法) 混合重み係数の値をゼロにして、平均値、分散値を所定の値にする。

(第 3 の方法) 混合重み係数の値を所定の値にして、平均値、分散値を標準モデルの出力分布を 1 個の分布に表現したときの平均値、分散値に

10 する。

のいずれかを利用して統計量の値を決定する。なお、利用する方法は、繰り返し回数 R、HMM、HMM の状態ごとに異なってもよい。ここでは、第 1 の方法を用いる。

統計量推定部 306 d は、このように推定した標準モデルの統計量を
15 統計量記憶部 306 c に記憶する。そして、このような統計量の推定と統計量記憶部 306 c への記憶を R (≧ 1) 回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル 322 の統計量として出力する。

次に、本実施の形態を PDA による環境音識別に適用した具体例を説
20 明する。

まず、参照モデル準備部 302 は、CD-ROM から環境音の識別に必要な参照モデルを読み出す。利用者は、識別を行う環境 (利用情報) を考慮して、識別したい環境音を画面上から選択する。たとえば、「乗用車」を選択し、続いて、「警報音」、「赤ちゃんの声」、「電車の音」などを

選択する。この選択に基づいて、参照モデル選択部 305 は、参照モデル記憶部 303 に記憶されている参照モデルの中から対応する参照モデルを選択する。そして、選択した参照モデル 323 を 1 つずつ用いて、標準モデル作成部 306 は、それぞれに対して標準モデルを作成する。

- 5 続いて、利用者は、PDA 301 において、「らくらく情報提供」（環境音に基づく状況判断による情報提供）というアプリケーションプログラムを起動する。このアプリケーションは、環境音に基づいて状況判断を行い、利用者に適切な情報を提供するプログラムである。起動されると、PDA 301 の表示画面に「正確に判断」、「素早く判断」という表示がされる。これに対して、利用者はどちらかを選択する。

- そして、仕様情報作成部 307 は、その選択結果に基づいて、仕様情報を作成する。たとえば、「正確に判断」が選択された場合には、精度を高くするために、混合分布数を 10 個とする仕様情報を作成する。一方、
15 「素早く判断」が選択された場合には、高速に処理するために、混合分布数を 1 個とする仕様情報を作成する。なお、複数の PDA が連携して処理できる場合などには、現在利用できる CPU パワーを判断し、その CPU パワーに基づいて仕様情報を作成してもよい。

- このような仕様情報にしたがって、「乗用車」、「警報音」、「赤ちゃんの声」、「電車の音」などの 1 混合の標準モデルが作成される。そして、P
20 DA 301 は、作成された標準モデルにより環境識別を行い、その識別結果に基づき、各種情報を PDA の画面に表示する。例えば、「乗用車」が近くにあると識別した場合は、道路地図を表示したり、「赤ちゃんの声」を識別した場合は、おもちゃ屋さんの広告を表示したりする。このようにして、本実施の形態における標準モデル作成装置によって作成された
25 標準モデルを用いた、環境音識別に基づく情報提供が実現される。なお、アプリケーションの仕様に応じて標準モデルの複雑さを調節することが

できる。

以上説明したように、本発明の第３の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成
5 されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

なお、統計量推定部 306d による処理の繰り返し回数は、上記式 2
10 5 に示された尤度の大きさがある一定のしきい値以上になるまでの回数としてもよい。

また、標準モデル 322 を構成する GMM は、雑音の種類ごとに異なる混合分布数の混合ガウス分布により構成してもよい。

また、識別モデルは、雑音モデルに限らず、話者を識別してもよいし、
15 年齢などを識別してもよい。

また、標準モデル 322 を CD-ROM、DVD-RAM、ハードディスクなどのストレージデバイスに記録してもよい。

また、参照モデル 321 を CD-ROM などのストレージデバイスから読み込む代わりに、PDA 301 において雑音データから参照モデル
20 321 を作成してもよい。

また、参照モデル準備部 302 は、必要に応じて CD-ROM などのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部 303 に追加・更新してもよいし、参照モデル記憶部 303 に格納されている不要な参照モデルを削除してもよい。

25 また、参照モデル準備部 302 は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部 303 に追加・更新してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、標準モデルの構造や、状態数などを決定してもよい。

- 5 また、近傍指示パラメータ G は、対象とする事象や標準モデルの出力分布によって異なってもよいし、繰り返し回数 R によって変化させてもよい。

（第 4 の実施の形態）

- 10 図 19 は、本発明の第 4 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ 401 に組み込まれた例が示されている。本実施の形態では顔認識用の標準モデルを作成する場合を例にして説明する。

- 15 サーバ 401 は、通信システムにおけるコンピュータ装置等であり、事象の出力確率によって定義される顔認識用の標準モデルを作成する標準モデル作成装置として、カメラ 411 と、画像データ蓄積部 412 と、参照モデル準備部 402 と、参照モデル記憶部 403 と、利用情報受信部 404 と、参照モデル選択部 405 と、標準モデル作成部 406 と、書き込み部 413 とを備える。

- 20 カメラ 411 により、顔の画像データが収集され、画像データ蓄積部 412 に顔画像データが蓄積される。参照モデル準備部 402 は、画像データ蓄積部 412 が蓄積した顔画像データを用いて話者ごとに参照モデル 421 を作成し、参照モデル記憶部 403 に記憶する。

- 25 利用情報受信部 404 は、利用者が希望する顔認識の対象となる人間の年齢の年代と性別の情報を利用情報 424 として電話 414 により受信する。参照モデル選択部 405 は、利用情報受信部 404 が受信した

利用情報 4 2 4 に基づいて、参照モデル記憶部 4 0 3 が記憶している参照モデル 4 2 1 の中から、利用情報 4 2 4 が示す年代と性別の話者に対応する参照モデル 4 2 3 を選択する。

標準モデル作成部 4 0 6 は、参照モデル選択部 4 0 5 が選択した話者の顔画像の参照モデル 4 2 3 に対する確率又は尤度を最大化又は極大化するように標準モデル 4 2 2 を作成する処理部であり、第 2 の実施の形態における標準モデル作成部 2 0 6 と同一の機能を有するとともに、第 1 の実施の形態における第 1 近似部 1 0 4 e と第 3 の実施の形態における第 2 近似部 3 0 6 e の機能を有する。つまり、第 1 ～ 第 3 の実施の形態で示された 3 種類の近似計算を組み合わせた計算を行う。

書き込み部 4 1 3 は、標準モデル作成部 4 0 6 が作成した標準モデル 4 2 2 を C D - R O M などのストレージデバイスに書き込む。

次に、以上のように構成されたサーバ 4 0 1 の動作について説明する。

図 2 0 は、サーバ 4 0 1 の動作手順を示すフローチャートである。図 2 1 は、サーバ 4 0 1 の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図 2 0 のステップ S 4 0 0）。つまり、カメラ 4 1 1 により A さんから Z さんの顔画像データを収集して画像データ蓄積部 4 1 2 に蓄積する。参照モデル準備部 4 0 2 は、画像データ蓄積部 4 1 2 が蓄積した顔画像データを用いて、話者ごとの参照モデル 4 2 1 を E M アルゴリズムにより作成する。ここでは参照モデル 4 2 1 は G M M で構成される。

参照モデル記憶部 4 0 3 は、参照モデル準備部 4 0 2 が作成した参照モデル 4 2 1 を記憶する。ここでは、図 2 1 の参照モデル 4 2 1 に示されるように、A さんから Z さんの全ての参照モデルが、混合分布数が 5 個の G M M により構成される。特徴量として 1 0 0 次元 ($J = 1 0 0$)

の画素の濃度値を用いる。

次に、利用情報受信部 404 は、利用情報 424 である年代と性別の情報を電話 414 により受信する（図 20 のステップ S401）。ここでは、利用情報 424 として、11 歳から 15 歳の男性と 22 歳から 26 歳の女性である。参照モデル選択部 405 は、その利用情報 424 に基づいて、参照モデル記憶部 403 が記憶している参照モデル 421 から、利用情報 424 に対応する参照モデル 423 を選択する（図 20 のステップ S402）。具体的には、図 21 の「選択された参照モデル 423」に示されるように、ここでは、11 歳から 15 歳の男性及び 22 歳から 26 歳の女性の参照モデルを選択する。

そして、標準モデル作成部 406 は、参照モデル選択部 405 が選択した話者の参照モデル 423 に対する確率又は尤度を最大化又は極大化するように標準モデル 422 を作成する（図 20 のステップ S403）。ここでは、図 21 の標準モデル 422 に示されるように、2 つの標準モデル 422 それぞれを、混合分布数が 3 個の GMM により構成する。

標準モデル 422 の作成方法は、基本的には、第 2 の実施の形態と同様に行われる。ただし、標準モデル 422 の統計量の推定における近似計算については、具体的には、以下のようにして行われる。つまり、標準モデル作成部 406 は、内蔵の記憶部等を介することで、第 1 の実施の形態における第 1 近似部 104 e による近似計算と同様の近似計算によって作成したモデルを初期値として、第 2 の実施の形態における一般近似部 206 e による近似計算と同様の近似計算による計算を行い、その結果を初期値として第 3 の実施の形態における第 2 近似部 306 e による近似計算と同様の近似計算を行う。

書き込み部 413 は、標準モデル作成部 406 が作成した 2 つの標準モデル 422 を CD-ROM などのストレージデバイスに書き込む（図

20のステップS404)。

利用者は、11歳から15歳の男性の標準モデルと22歳から26歳の女性の標準モデルが書き込まれたストレージデバイスを郵送で受け取る。

- 5 次に、本実施の形態を、行動予測に基づいてお店などを紹介する情報提供システムに適用した具体例を説明する。この情報提供システムは、通信ネットワークで接続されたカーナビゲーション装置と情報提供サーバ装置から構成される。カーナビゲーション装置は、本実施の形態における標準モデル作成装置401によって予め作成された標準モデルを行
- 10 動予測モデルとして利用することで、人の行動（つまり、車による行先等）を予測し、その行動に関連した情報（行先の近くに位置するレストランなどのお店の情報など）を提供する機能を備える。

- まず、利用者は、カーナビゲーション装置を用いて、電話回線414で接続されたサーバ401に対して、自分用の行動予測モデルの作成を
- 15 依頼する。

具体的には、利用者は、カーナビゲーション装置が表示する項目選択画面で、「らくらく推薦機能」のボタンを押す。すると、利用者の住所（利用場所）、年齢、性別、趣味などを入力する画面になる。

- ここでは、利用者はお父さんとお母さんとする。まず、お父さんの個人情報をカーナビゲーション装置の画面と対話しながら入力する。住所
- 20 については、電話番号を入力することにより自動的に変換される。あるいは、カーナビゲーション装置において現在位置が表示されているときに「利用場所」のボタンを押すことで、その現在位置が利用場所として入力される。ここでは住所の情報を住所Aとする。年齢と性別については、
- 25 「50代」、「男」を選択して入力する。趣味については、予め表示されたチェック項目があるので、利用者は、該当箇所をチェックする。こ

ここではお父さんの趣味の情報を趣味情報Aとする。

続いて、お母さんの個人情報についても同様に入力する。住所B、40代、女、趣味情報Bからなる個人情報が作成される。このような入力の結果は、図22の画面表示例に示されるとおりである。

- 5 最後に、カーナビゲーション装置は、このようにして作成された個人情報を利用情報として、付属の電話回線414を用いて、情報提供サーバ装置であるサーバ401に転送する。

- 次に、サーバ401は、転送されてきた個人情報（利用情報）に基づいて、お父さんとお母さんの2個の行動予測モデルを作成する。ここで、
10 行動予測モデルは、確率モデルで表現され、その入力は、曜日、時刻、現在地などで、出力は、お店Aの情報を提示する確率、お店Bの情報を提示する確率、お店Cの情報を提示する確率、駐車場の情報を提示する確率などである。

- サーバ401の参照モデル記憶部403に記憶されている複数の参照
15 モデルは、年代、性別、代表的な住所と趣味の傾向で作成した行動予測モデルである。サーバ401では、予め、カメラ411に代えて、カーナビゲーション装置の入力ボタン等を用いて各種個人情報（上記入力及び出力についての情報）を入力することで、画像データ蓄積部412に各種個人情報を蓄積したうえで、参照モデル準備部402によって、
20 画像データ蓄積部412に蓄積された個人情報から、複数種類の典型的な利用者ごとの参照モデル421が作成され、参照モデル記憶部403に格納されている。

- 参照モデル選択部405は、個人情報（利用情報）を用いて、個人情報にふさわしい参照モデルを選択する。例えば、同じ町の、年代と性別
25 が同じで、趣味のチェック項目が8割以上一致した参照モデルを選択する。サーバ401の標準モデル作成部406は、選択された参照モデル

を統合した標準モデルを作成する。作成された標準モデルは書き込み部 413 により、メモリカードに記憶される。ここでは、お父さんとお母さんの 2 人の標準モデルが記憶される。メモリカードは、郵送で利用者に届けられる。

- 5 利用者は、受け取ったメモリカードをカーナビゲーション装置に挿入し、画面に表示された「お父さん」と「お母さん」を選択することで、利用者を設定する。これによって、カーナビゲーション装置は、装着されたメモリカードに記憶された標準モデルを行動予測モデルとして使用することで、現在の曜日、時刻、場所などから、必要なタイミングでお店 10 の情報などを提示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを行動予測モデルとして用いることで、人の行動（つまり、車による行先）を予測し、その行動に関連した情報を提供する情報提供システムが実現される。

- 15 以上説明したように、本発明の第 4 の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい高精度な標準モデルが提供される。

- 20 なお、標準モデル 422 を構成する GMM は、話者ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル準備部 402 は、必要に応じて新たな参照モデルを作成して参照モデル記憶部 403 に追加・更新してもよいし、参照モデル記憶部 403 に格納されている不要な参照モデルを削除してもよい。

- 25 また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、標準モデルの構造や、状態数などを

決定してもよい。

(第 5 の実施の形態)

図 2 3 は、本発明の第 5 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ 5 0 1 に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

サーバ 5 0 1 は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部 5 1 1 と、音声データ蓄積部 5 1 2 と、参照モデル準備部 5 0 2 と、参照モデル記憶部 5 0 3 と、利用情報受信部 5 0 4 と、参照モデル選択部 5 0 5 と、標準モデル作成部 5 0 6 と、仕様情報受信部 5 0 7 と、書き込み部 5 1 3 とを備える。

読み込み部 5 1 1 は、C D - R O M などのストレージデバイスに書き込まれた子供、成人、高齢者の音声データを読み込み、音声データ蓄積部 5 1 2 に蓄積する。参照モデル準備部 5 0 2 は、音声データ蓄積部 5 1 2 が蓄積した音声データを用いて話者ごとに参照モデル 5 2 1 を作成する。参照モデル記憶部 5 0 3 は、参照モデル準備部 5 0 2 が作成した参照モデル 5 2 1 を記憶する。

仕様情報受信部 5 0 7 は、仕様情報 5 2 5 を受信する。利用情報受信部 5 0 4 は、利用情報 5 2 4 である利用者の音声を受信する。参照モデル選択部 5 0 5 は、利用情報 5 2 4 である利用者の音声に音響的に近い話者の参照モデルを、参照モデル記憶部 5 0 3 が記憶している参照モデル 5 2 1 から選択する。

標準モデル作成部 5 0 6 は、仕様情報 5 2 5 に基づいて、参照モデル

選択部 505 が選択した話者の参照モデル 523 に対する確率又は尤度を最大化又は極大化するように標準モデル 522 を作成する処理部であり、第 1 の実施の形態における標準モデル作成部 104 と同一の機能を有する。書き込み部 513 は、標準モデル作成部 506 が作成した標準モデル 522 を CD-ROM などのストレージデバイスに書き込む。

次に、以上のように構成されたサーバ 501 の動作について説明する。

図 24 は、サーバ 501 の動作手順を示すフローチャートである。図 25 は、サーバ 501 の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

10 まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図 24 のステップ S500）。つまり、読み込み部 511 は、CD-ROM などのストレージデバイスに書き込まれた音声データを読み込み、音声データ蓄積部 512 に蓄積する。参照モデル準備部 502 は、音声データ蓄積部 512 が蓄積した音声データを用いて話者ごとに参照
15 モデル 521 をバウム・ウェルチの再推定の方法により作成する。参照モデル記憶部 503 は、参照モデル準備部 502 が作成した参照モデル 521 を記憶する。

参照モデル 521 は、音素ごとの HMM により構成される。ここでは、図 25 の参照モデル 521 に示されるように、子供の各話者の参照モデルは、状態数 3 個、各状態は混合分布数が 3 個の混合ガウス分布により
20 HMM の出力分布が構成され、成人の各話者の参照モデルが、状態数 3 個、各状態は混合分布数が 6 4 個の混合ガウス分布により HMM の出力分布が構成され、高齢者の各話者の参照モデルは、状態数 3 個、各状態は混合分布数が 1 6 個の混合ガウス分布により HMM の出力分布が構成
25 される。これは、子供の音声データが比較的少なく、成人の音声データが多いためである。特徴量として 25 次元（ $J=25$ ）のメルケプスト

ラム係数が用いられる。

次に、利用情報受信部 504 は、利用者の音声、端末装置 514 から、利用情報 524 として受信する（図 24 のステップ S501）。参照モデル選択部 505 は、利用情報 524 である利用者の音声に音響的に
5 近い参照モデル 523 を、参照モデル記憶部 503 が記憶している参照モデル 521 から選択する（図 24 のステップ S502）。具体的には、図 25 の「選択された参照モデル 523」に示されるように、ここでは、近い話者 10 人（ $N_g = 10$ ）の参照モデルが選択される

そして、仕様情報受信部 507 は、利用者の要求に基づき仕様情報 5
10 25 を端末装置 514 から受信する（図 24 のステップ S503）。ここでは、速い認識処理という仕様情報 525 を受信する。標準モデル作成部 506 は、仕様情報受信部 507 が受信した仕様情報 525 に基づいて、参照モデル選択部 505 が選択した話者の参照モデル 523 に対する確率又は尤度を最大化又は極大化するように標準モデル 522 を作成
15 する（図 24 のステップ S504）。具体的には、標準モデル 522 は、図 25 の標準モデル 522 に示されるように、仕様情報 525 である速い認識処理という情報に基づいて、2 混合（ $M_f = 2$ ）で、3 状態の HMM より構成する。HMM は音素ごとに構成する。

標準モデル 522 の作成方法は、第 1 の実施の形態と同様に行われる。
20 書き込み部 513 は、標準モデル作成部 506 が作成した標準モデル 522 を CD-ROM などのストレージデバイスに書き込む（図 24 のステップ S505）。

次に、本実施の形態を、通信ネットワークを用いた音声認識によるゲームに適用した具体例を説明する。ここでは、サーバ 501 は、作成した標準モデルを用いて音声認識を行う音声認識部を備えるものとする。
25 また、端末装置 514 として、PDA とする。これらは、通信ネットワ

ークで接続されている。

サーバ501では、読み込み部511、音声データ蓄積部512及び参照モデル準備部502により、音声データをCDやDVDなどで入手したタイミングで参照モデルを逐次準備している。

- 5 利用者は、PDA（端末装置514）において、音声認識を利用したゲームプログラム、ここでは、「アクションゲーム」を立ち上げる。すると、『アクション』と発声してください』と表示されるので、利用者は、「アクション」と発声する。その音声は、利用情報として、PDA（端末装置514）からサーバ501に送信され、サーバ501の利用情報
10 受信部504及び参照モデル選択部505により、参照モデル記憶部503に記憶された複数の参照モデルの中から利用者に合った参照モデルを選択する。

- また、利用者は、速くリアクションしてほしいので、PDA（端末装置514）の設定画面において「高速に認識する」と設定する。その設
15 定内容は、仕様情報として、PDA（端末装置514）からサーバ501に送信され、サーバ501においては、このような仕様情報及び選択された参照モデルに基づいて、標準モデル作成部506により、2混合の標準モデルが作成される。

- 利用者は、アクションゲームにおいて、PDAのマイクに「右に移動」、
20 「左に移動」などのコマンドを発声する。入力された音声は、サーバへ送信され、既に作成された標準モデルを利用した音声認識が行われる。その認識結果は、サーバ501からPDA（端末装置514）に送信され、PDA（端末装置514）において、送信されてきた認識結果に基づいて、アクションゲームのキャラクタが動く。このようにして、本実
25 施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声によるアクションゲームが実現される。

また、同様にして、本実施の形態を別のアプリケーション、例えば、通信ネットワークを用いた翻訳システムに適用することもできる。たとえば、利用者は、PDA（端末装置 514）において、「音声翻訳」というアプリケーションプログラムを立ち上げる。すると、『翻訳』と発声
5 してください」と表示される。利用者は、「翻訳」と発声する。その音声は、利用情報として、PDA（端末装置 514）からサーバ 501 に送信される。また、利用者は、正確に認識してほしいので、そのアプリケーションにおいて、「正確に認識してほしい」旨を指示する。その指示は、仕様情報として、PDA（端末装置 514）からサーバ 501 に送信さ
10 れる。サーバ 501 では、送信されてきた利用情報及び仕様情報に従って、たとえば、100 混合の標準モデルが作成される。

利用者は、PDA（端末装置 514）のマイクに向かって「おはようございます」と発声する。入力された音声は PDA（端末装置 514）からサーバ 501 に送信され、サーバ 501 で「おはようございます」と認識された後に、その認識結果が PDA（端末装置 514）に返信さ
15 れる。PDA（端末装置 514）は、サーバ 501 から受信した認識結果を英語に翻訳し、その結果「GOOD MORNING」を画面に表示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声による翻
20 訳装置が実現される。

以上説明したように、本発明の第 5 の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供さ
25 れる。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデル

を利用する機器にふさわしい標準モデルが準備される。

また、参照モデル準備部 502 において、参照モデルごとにデータ数に適した混合分布数の精度の高い参照モデルを準備でき、精度の高い参照モデルを用いて標準モデルを作成できる。このため精度の高い標準モデルの利用が可能となる。

なお、標準モデル 522 は、音素ごとに HMM を構成するに限らず、文脈依存の HMM で構成してもよい。

また、標準モデル 522 を構成する HMM は、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、標準モデル 522 を用いて、サーバ 501 において音声認識を行ってもよい。

また、参照モデル準備部 502 は、必要に応じて新たな参照モデルを作成して参照モデル記憶部 503 に追加・更新してもよいし、参照モデル記憶部 503 に格納されている不要な参照モデルを削除してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、標準モデルの構造や、状態数などを決定してもよい。

(第 6 の実施の形態)

図 26 は、本発明の第 6 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ 601 に組み込まれた例が示されている。本実施の形態では意図理解のための標準モデル（嗜好モデル）を作成する場合を例にして説明する。

サーバ 601 は、通信システムにおけるコンピュータ装置等であり、事象の出力確率によって定義される意図理解用の標準モデルを作成する

標準モデル作成装置として、読み込み部 611 と、参照モデル準備部 602 と、参照モデル記憶部 603 と、利用情報受信部 604 と、参照モデル選択部 605 と、標準モデル作成部 606 と、仕様情報作成部 607 とを備える。

5 読み込み部 611 は、CD-ROM などのストレージデバイスに書き込まれた年齢別の話者 A さんから話者 Z さんの嗜好モデルを読み込み、参照モデル準備部 602 は、読み込まれた参照モデル 621 を参照モデル記憶部 603 へ送信し、参照モデル記憶部 603 は、参照モデル 621 を記憶する。

10 仕様情報作成部 607 は、普及しているコンピュータの CPU パワーに合わせて仕様情報 625 を作成する。利用情報受信部 604 は、端末装置 614 から利用情報 624 を受信する。参照モデル選択部 605 は、利用情報受信部 604 が受信した利用情報 624 に基づいて、参照モデル記憶部 603 が記憶している参照モデル 621 からの中から、利用情報 15 624 に対応した参照モデル 623 を選択する。

標準モデル作成部 606 は、仕様情報作成部 607 が作成した仕様情報 625 に基づいて、参照モデル選択部 605 が選択した参照モデル 623 に対する確率又は尤度を最大化又は極大化するように標準モデル 622 を作成する処理部であり、第 2 の実施の形態における標準モデル作成部 206 と同一の機能を有するとともに、第 3 の実施の形態における第 2 近似部 306 e の機能を有する。つまり、第 2 及び第 3 の実施の形態で示された 2 種類の近似計算を組み合わせた計算を行う。

次に、以上のように構成されたサーバ 601 の動作について説明する。

図 27 は、サーバ 601 の動作手順を示すフローチャートである。図 28 は、サーバ 601 の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図 27 のステップ S 6 0 0）。つまり、読み込み部 6 1 1 は、CD-ROM などのストレージデバイスに書き込まれた年齢別の話者 A さんから話者 Z さんの嗜好モデルを読み込み、参照モデル準備部 6 0 2 は、読み込まれた参照モデル 6 2 1 を参照モデル記憶部 6 0 3 へ送信し、参照モデル記憶部 6 0 3 は、参照モデル 6 2 1 を記憶する。

参照モデル 6 2 1 は、GMM より構成される。ここでは、図 28 の参照モデル 6 2 1 に示されるように、混合分布数が 3 個の GMM より構成される。学習データとして、趣味、性格などを数値化した 5 次元（ $J = 5$ ）の特徴量を用いる。参照モデルの準備は、標準モデルの作成が要求される以前に行う。

次に、利用情報受信部 6 0 4 は、嗜好モデルを作成したい年齢層である利用情報 6 2 4 を受信する（図 27 のステップ S 6 0 1）。ここでは、20 代、30 代、40 代の年代別の嗜好モデルを利用するという利用情報 6 2 4 である。参照モデル選択部 6 0 5 は、図 28 の「選択された参照モデル 6 2 3」に示されるように、利用情報受信部 6 0 4 が受信した利用情報 6 2 4 が示す年代の話者の嗜好モデルを、参照モデル記憶部 6 0 3 が記憶している参照モデル 6 2 1 から選択する（図 27 のステップ S 6 0 2）。

そして、仕様情報作成部 6 0 7 は、普及しているコンピュータの CPU パワー、記憶容量などに基づき仕様情報 6 2 5 を作成する（図 27 のステップ S 6 0 3）。ここでは、通常速度の認識処理という仕様情報 6 2 5 を作成する。

標準モデル作成部 6 0 6 は、仕様情報作成部 6 0 7 が作成した仕様情報 6 2 5 に基づいて、参照モデル選択部 6 0 5 が選択した話者の参照モデル 6 2 3 に対する確率又は尤度を最大化又は極大化するように標準モ

デル 6 2 2 を作成する（図 2 7 のステップ S 6 0 4）。ここでは、標準モデル 6 2 2 は、図 2 8 の標準モデル 6 2 2 に示されるように、仕様情報 6 2 5 である通常速度の認識処理という情報に基づいて 3 混合（ $Mf=3$ ）の GMM より構成する。

- 5 標準モデル 6 2 2 の作成方法は、基本的には、第 2 の実施の形態と同様に行われる。ただし、標準モデル 6 2 2 の統計量の推定における近似計算については、具体体には、以下のようにして行われる。つまり、標準モデル作成部 6 0 6 は、内蔵の記憶部等を介することで、第 2 の実施の形態における一般近似部 2 0 6 e による近似計算と同様の近似計算に
10 による計算を行い、その結果を初期値として第 3 の実施の形態における第 2 近似部 3 0 6 e による近似計算と同様の近似計算を行う。

次に、本実施の形態を情報検索装置に適用した具体例を説明する。ここでは、参照モデルは、入力が検索キーワードであり、出力が検索ルール A、検索ルール B などを利用する確率である。異なる検索ルールを用
15 いると、表示される検索結果が異なってくる。また、サーバ 6 0 1 の参照モデル記憶部 6 0 3 に準備される参照モデルは、代表的な特徴をもつ話者のモデルとする。

まず、利用者は、サーバ 6 0 1 に付属しているリモコン（端末装置 6 1 4）を用いて利用情報を入力する。利用情報は、年齢、性格、性別、
20 趣味などでである。また、「子供」、「俳優」、「高校生」などの所定のグループを識別する情報であってもよい。

続いて、利用者は、選択画面で、「カーナビゲーション装置用」、「携帯電話用」、「パソコン用」、「テレビ用」などから 1 つの利用機器を選択する。サーバ 6 0 1 の仕様情報作成部 6 0 7 は、利用機器の CPU パワー、
25 記憶容量に基づいて仕様情報を作成する。ここでは、「テレビ用」が選択されたとし、CPU パワーと記憶容量が小さい旨の仕様情報 6 2 5 が作

成され、その仕様情報 6 2 5 に基づいて、標準モデル作成部 6 0 6 によって、小さい CPU パワーでも動作する 3 混合の標準モデルが作成される。作成された標準モデルはメモリカードに書き込まれ、そのメモリカードは利用者によってテレビに挿入される。

- 5 利用者は、テレビに表示された E P G など、おすすめ番組を検索するために検索キーワードを入力する。すると、テレビは、メモリカードに記録された標準モデルを用いて、検索キーワードに合った検索ルールを決定し、その検索ルールに沿って番組を検索し、利用者の嗜好にあった番組として表示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた便利な検索装置が
- 10 実現される。

- 以上説明したように、本発明の第 6 の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成
- 15 されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

- なお、標準モデル 6 2 2 を構成する G M M は、話者ごとに異なる分布
- 20 数の混合ガウス分布により構成してもよい。

また、参照モデル準備部 6 0 2 は、必要に応じて C D - R O M などのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部 6 0 3 に追加・更新してもよいし、参照モデル記憶部 6 0 3 に格納されている不要な参照モデルを削除してもよい。

- 25 また、参照モデル及び標準モデルの G M M はベイジアンネットの一部を表現するものでもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

5 (第7の実施の形態)

図29は、本発明の第7の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ701に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

サーバ701は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部711と、参照モデル準備部702と、参照モデル記憶部703と、
15 利用情報受信部704と、参照モデル選択部705と、標準モデル作成部706と、仕様情報受信部707と、標準モデル記憶部708と、標準モデル送信部709とを備える。

参照モデル準備部702は、読み込み部711が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部703へ送信し、参照モデル記憶部703は、送信された参照モデル721を記憶する。

仕様情報受信部707は、端末装置712から仕様情報725を受信する。利用情報受信部704は、端末装置712から利用情報724である雑音下で発声した利用者の音声を受信する。参照モデル選択部705は、利用情報724である利用者の音声に音響的に近い話者・雑音・声調子の参照モデル723を、参照モデル記憶部703が記憶している

参照モデル 7 2 1 の中から選択する。

標準モデル作成部 7 0 6 は、仕様情報受信部 7 0 7 が受信した仕様情報 7 2 5 に基づいて、参照モデル選択部 7 0 5 が選択した参照モデル 7 2 3 に対する確率又は尤度を最大化又は極大化するように標準モデル 7 2 2 を作成する処理部であり、第 2 の実施の形態における標準モデル作成部 2 0 6 と同一の機能を有する。標準モデル記憶部 7 0 8 は、仕様情報 7 2 5 に基づいた 1 もしくは複数の標準モデルを記憶する。標準モデル送信部 7 0 9 は、利用者の端末装置 7 1 2 から仕様情報と標準モデルの要求信号を受信すると、その仕様情報に適した標準モデルを端末装置 7 1 2 へ送信する。

次に、以上のように構成されたサーバ 7 0 1 の動作について説明する。

図 3 0 は、サーバ 7 0 1 の動作手順を示すフローチャートである。図 3 1 は、サーバ 7 0 1 の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する(図 3 0 のステップ S 7 0 0)。つまり、参照モデル準備部 7 0 2 は、読み込み部 7 1 1 が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部 7 0 3 へ送信し、参照モデル記憶部 7 0 3 は、送信された参照モデル 7 2 1 を記憶する。ここでは、参照モデル 7 2 1 は、話者・雑音・声の調子ごとに、音素ごとの HMM により構成される。また、各参照モデルは、図 3 1 の参照モデル 7 2 1 に示されるように、状態数 3 個、各状態は混合分布数が 1 2 8 個の混合ガウス分布により HMM の出力分布が構成される。特徴量として 2 5 次元 ($J = 25$) のメルケプストラム係数が用いられる。

次に、利用情報受信部 7 0 4 は、利用者 A の雑音下での音声を端末装

置 7 1 2 から利用情報 7 2 4 として受信する（図 3 0 のステップ S 7 0 1）。参照モデル選択部 7 0 5 は、利用情報 7 2 4 である利用者 A の音声に音響的に近い参照モデル 7 2 3 を、参照モデル記憶部 7 0 3 が記憶している参照モデル 7 2 1 の中から選択する（図 3 0 のステップ S 7 0 2）。

- 5 具体的には、図 3 1 の「選択された参照モデル 7 2 3」に示されるように、ここでは、近い話者 1 0 0 人（ $N_g = 1 0 0$ ）の参照モデルが選択される

- そして、仕様情報受信部 7 0 7 は、利用者 A の要求に基づき仕様情報 7 2 5 を端末装置 7 1 2 から受信する（図 3 0 のステップ S 7 0 3）。
10 ここでは、高い認識精度という仕様情報 7 2 5 を受信する。標準モデル作成部 7 0 6 は、仕様情報 7 2 5 に基づいて、参照モデル選択部 7 0 5 が選択した参照モデル 7 2 3 に対する確率又は尤度を最大化又は極大化するように標準モデル 7 2 2 を作成する（図 3 0 のステップ S 7 0 4）。
15 具体的には、標準モデル 7 2 2 は、図 3 1 の標準モデル 7 2 2 に示されるように、仕様情報 7 2 5 である高い認識精度という情報に基づいて、6 4 混合（ $M_f = 6 4$ ）で、3 状態の HMM より構成する。HMM は音素ごとに構成する。

標準モデル 7 2 2 の作成方法は、第 2 の実施の形態と同様に行われる。

- 標準モデル記憶部 7 0 8 は、仕様情報 7 2 5 に基づいた 1 もしくは複
20 数の標準モデル 7 2 2 を記憶する。ここでは、以前に作成した標準モデルである利用者 B の 1 6 混合の HMM がすでに記憶されており、新たに利用者 A の 6 4 混合の HMM が記憶される。

- 利用者 A は、端末装置 7 1 2 からサーバ 7 0 1 の標準モデル送信部 7 0 9 へ、仕様情報である利用者 A と雑音の種類と標準モデルの要求信号
25 とを送信する（図 3 0 のステップ S 7 0 6）。標準モデル送信部 7 0 9 は、

その端末装置 7 1 2 へ、仕様に適した標準モデルを端末装置 7 1 2 へ送信する（図 3 0 のステップ S 7 0 7）。ここでは、先ほど作成した利用者 A の標準モデル 7 2 2 を端末装置 7 1 2 へ送信する。

利用者 A は端末装置 7 1 2 において受信した標準モデル 7 2 2 を用いて音声認識を行う（図 3 0 のステップ S 7 0 8）。

次に、本実施の形態を、通信ネットワークで接続されたカーナビゲーション装置（端末装置 7 1 2）とサーバ装置（サーバ 7 0 1；標準モデル作成装置）から構成される音声認識システムに適用した具体例を説明する。

10 まず、利用者は、カーナビゲーション装置（端末装置 7 1 2）の画面にて「自分の音声モデルを獲得」する旨のボタンを選択する。すると、「名前を入力」と表示されるので、ボタン操作により自分の名前を入力する。次に、『音声』と発声してください」と表示されるので、利用者は、カーナビゲーション装置付属のマイクに向かって「音声」と発声する。これらの情報（利用者の名前、雑音下での音声）は、利用情報として、カーナビゲーション装置（端末装置 7 1 2）からサーバ 7 0 1 に送信される。

同様に、利用者は、カーナビゲーション装置（端末装置 7 1 2）の画面にて「高精度の音声認識」のボタンを選択する。すると、その選択情報は、仕様情報として、カーナビゲーション装置（端末装置 7 1 2）からサーバ 7 0 1 に送信される。

サーバ 7 0 1 は、それらの利用情報及び仕様情報に基づいて、利用者 にふさわしい音声認識用の標準モデルを作成し、作成した標準モデルを利用者の名前と対応づけて標準モデル記憶部 7 0 8 に格納しておく。

25 次回にカーナビゲーション装置（端末装置 7 1 2）を起動すると、「名

の名前がサーバ 701 に送信され、標準モデル 722 に格納された対応する標準モデルが標準モデル送信部 709 によってサーバ 701 から端末装置 712 に送信される。名前（利用者）に対応した標準モデルをサーバ 701 からダウンロードした端末装置 712 は、その標準モデルを用いて、利用者に対する音声認識を行い、音声による目的地設定などを行う。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声によってカーナビゲーション装置を操作することが可能となる。

以上説明したように、本発明の第 7 の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

また、標準モデル記憶部 708 は、複数の標準モデルを記憶することができるため、必要に応じてすぐに標準モデルが提供される。

また、標準モデル送信部 709 により、標準モデルが端末装置 712 へ送信されるので、端末装置 712 とサーバ 701 が空間的に離れた場所に設置してある場合に、端末装置 712 は、容易にサーバ 701 が作成した標準モデルを利用することができる。

なお、標準モデル 722 は、音素ごとに HMM を構成するに限らず、文脈依存の HMM で構成してもよい。

また、標準モデル 722 を構成する HMM は、状態ごとに異なる混合数の混合ガウス分布により構成してもよい。

また、標準モデル 722 を用いて、サーバ 701 において音声認識を

行い、認識結果を端末装置 712 へ送信してもよい。

また、参照モデル準備部 702 は、必要に応じて新たな参照モデルを作成して参照モデル記憶部 703 に追加・更新してもよいし、参照モデル記憶部 703 に格納されている不要な参照モデルを削除してもよい。

- 5 また、参照モデル準備部 702 は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部 703 に追加・更新してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

- 10 また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などの HMM の構造や、状態数などを決定してもよい。

(第 8 の実施の形態)

- 図 32 は、本発明の第 8 の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置が携帯電話 901 に組み込まれた例が示されている。本実施の
15 形態では音声認識用の標準モデルを作成する場合を例にして説明する。

- 携帯電話 901 は、携帯情報端末であり、事象の集合と事象又は事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、参照モデル受信部 909 と、参照モデル準備部 902 と、参照モデル記憶部
20 903 と、利用情報作成部 904 と、参照モデル選択部 905 と、類似度情報作成部 908 と、標準モデル作成部 906 と、仕様情報作成部 907 と、マイク 912 と、音声認識部 913 とを備える。

利用情報作成部 904 は、利用情報 924 を携帯電話 901 の画面とキーを利用して作成する。

- 25 仕様情報作成部 907 は、携帯電話 901 の仕様に基づき仕様情報 925 を作成する。ここで、仕様情報とは、作成する標準モデルの仕様に

関する情報であり、ここでは、携帯電話 901 が備える CPU の処理能力に関する情報である。

類似度情報作成部 908 は、利用情報 924 と仕様情報 925 と参照モデル記憶部 903 が記憶した参照モデル 921 に基づいて、類似度情報 926 を作成して参照モデル準備部に送信する。

参照モデル準備部 902 は、類似度情報 926 に基づいて、参照モデルを準備するか否かを決定する。参照モデル準備部 902 は、参照モデルを準備すると決定した場合に、利用情報 924 と仕様情報 925 を参照モデル受信部 909 に送信する。

10 参照モデル受信部 909 は、利用情報 924 と仕様情報 925 に対応した参照モデルを、サーバ装置 910 から受信して参照モデル準備部 902 に送信する。

参照モデル準備部 902 は、参照モデル受信部 909 が送信した参照モデルを参照モデル記憶部 903 に記憶する。

15 参照モデル選択部 905 は、利用情報 924 に対応した参照モデル 923 を、参照モデル記憶部 903 が記憶している参照モデル 921 の中から選択する。

標準モデル作成部 906 は、仕様情報作成部 907 で作成された仕様情報 925 に基づいて、参照モデル選択部 905 が選択した参照モデル 923 に対する確率又は尤度を最大化又は極大化するように標準モデル 922 を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部 906a と、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部 906b と、決定された初期標準モデルを
20 記憶する統計量記憶部 906c と、統計量記憶部 906c に記憶された

ることにより、参照モデル選択部 905 が選択した参照モデル 923 に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部 906 d とからなる。

音声認識部 913 は、標準モデル作成部 906 で作成された標準モデル 922 を用いて、マイク 912 から入力された利用者の音声を認識する。

次に、以上のように構成された携帯電話 901 の動作について説明する。

図 33 は、携帯電話 901 の動作手順を示すフローチャートである。

いま、参照モデル記憶部 903 には、あらかじめ参照モデル 921 として子供用モデルが記憶されているとする。その参照モデル 921 は、音素ごとの HMM により構成される。参照モデル 921 の一例を図 34 に示す。ここでは、子供用参照モデルのイメージ図が示されている。これらの参照モデルは、状態数 3 個、各状態は分布数が 16 個の混合ガウス分布により HMM の出力分布が構成される。特徴量として、12 次元のメルケプストラム係数、12 次元のデルタメルケプストラム係数、デルタパワーの合計 25 次元 ($J = 25$) の特徴量が用いられる。

まず、利用情報作成部 904 は、利用者の属するカテゴリである利用情報 924 を作成する（ステップ S900）。図 36 は、利用情報 924 の作成例を示す図である。図 36 (a) に携帯電話 901 の選択画面の一例を示す。ここでは、「4：成人」のボタンを押すことにより、この携帯電話 901 が成人女性と成人男性に利用されることが選択されている。別の一例を図 36 (b) に示す。ここでは、「メニュー」ボタンを押しながら音声を入力している。その利用者の音声は、特徴量に変換されることで、利用情報 924 である「利用者の音声データ」が作成される。

一方、仕様情報作成部 907 は、携帯電話 901 の仕様に基づき、仕

様情報 925 を作成する (ステップ S901)。ここでは、携帯電話 901 のメモリ容量の大きさに基づいて「混合分布数 16」という仕様情報 925 を作成する。

次に、類似度情報作成部 908 は、利用情報 924 と仕様情報 925
5 と参照モデル記憶部 903 が記憶した参照モデル 921 に基づいて、類似度情報 926 を作成して (ステップ S902)、類似度情報 926 を参照モデル準備部 902 に送信する。ここでは、参照モデル記憶部 903 に存在する参照モデル 921 は、混合分布数 3 の子供用モデル (図 34 を参照) のみであり、利用情報 924 である「成人」(図 36 (a) に対応) と仕様情報 925 である「混合分布数 16」に対応する参照モデル
10 が参照モデル記憶部 903 に存在しないため、「類似した参照モデルが存在しない」という類似度情報 926 を作成して、類似度情報 926 を参照モデル準備部 902 に送信する。別の一例では、利用情報 924 は「利用者の音声データ」(図 36 (b) に対応) であり、利用者の音声データ
15 を参照モデル記憶部 903 が記憶している子供用モデルに入力して類似度情報 926 を作成する。ここでは、子供用モデルに対する尤度が所定のしきい値以下であるため、「類似した参照モデルが存在しない」という類似度情報 926 を作成して参照モデル準備部 902 に送信する。

続いて、参照モデル準備部 902 は、類似度情報 926 に基づいて、
20 参照モデルを準備するか否かを決定する (ステップ S903)。ここでは、「類似した参照モデルが存在しない」ため、図 37 (a) の携帯電話 901 の画面表示例に示すように利用者に参照モデルの準備を促す。ここで、利用者が「メモ」ボタンを押して参照モデルの準備を要求した場合に、参照モデル準備部 902 は、参照モデルを準備すると決定して、利用
25 情報 924 と仕様情報 925 を参照モデル受信部 909 に送信する。別の一例では、「類似した参照モデルが存在しない」ため、参照モデル準

備部 902 は、自動的に参照モデルを準備すると決定して、利用情報 924 と仕様情報 925 を参照モデル受信部 909 に送信する。この場合の携帯電話 901 の画面の一例を図 37 (b) に示す。

これに対して、参照モデル受信部 909 は、利用情報 924 と仕様情報 925 に対応した参照モデルをサーバ装置 910 から受信して参照モデル準備部 902 に送信する。ここでは、参照モデル受信部 909 は、利用情報 924 である「成人」(図 36 (a) に対応) と仕様情報 925 である「混合分布数 16」に対応する参照モデルである、「混合分布数 16 の成人女性用モデル」と「混合分布数 16 の成人男性用モデル」の 2 個の参照モデルをサーバ装置 910 から受信する。

そして、参照モデル準備部 902 は、参照モデル受信部 909 が送信した参照モデルを参照モデル記憶部 903 に記憶することによって参照モデルを準備する(ステップ S904)。図 35 にその参照モデルの一例を示す。ここでは、成人男性用、成人女性用、子供用の参照モデルのイメージ図が示されている。

次に、参照モデル選択部 905 は、利用情報 924 である「成人」に対応した同じカテゴリに属する「混合分布数 16 の成人女性用モデル」と「混合分布数 16 の成人男性用モデル」の 2 個の参照モデルを参照モデル記憶部 903 が記憶している参照モデル 921の中から選択する(ステップ S905)。別の一例では、参照モデル選択部 905 は、利用情報 924 である「利用者の音声データ」と音響的に近い(尤度が高い)「混合分布数 16 の成人女性用モデル」と「混合分布数 16 の成人男性用モデル」の 2 個の参照モデルを参照モデル記憶部 903 が記憶している参照モデル 921の中から選択する。

続いて、標準モデル作成部 906 は、作成された仕様情報 925 に基づいて、参照モデル選択部 905 が選択した参照モデル 923 に対する

確率又は尤度を最大化又は極大化するように標準モデル 9 2 2 を作成する（ステップ S 9 0 6）。

最後に、音声認識部 9 1 3 は、標準モデル作成部 9 0 6 によって作成された標準モデル 9 2 2 に従って、マイク 9 1 2 から入力された利用者
5 の音声認識する（ステップ S 9 0 7）。

次に、図 3 3 におけるステップ S 9 0 6（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図 4 に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

10 まず、標準モデル構造決定部 9 0 6 a は、標準モデルの構造を決定する（図 4 のステップ S 1 0 2）。ここでは、標準モデルの構造として、仕様情報 9 2 5 である「混合分布数 1 6」に基づいて、音素ごとの H M M により構成し、状態数を 3 とし、各状態における出力分布の混合分布数を 1 6 個（ $Mf = 16$ ）と決定する。

15 次に、初期標準モデル作成部 9 0 6 b は、標準モデルを計算するための統計量の初期値を決定する（図 4 のステップ S 1 0 2 b）。ここでは、選択された参照モデル 9 2 3 である「混合分布数 1 6 の成人女性用モデル」を統計量の初期値として統計量記憶部 9 0 6 c に記憶する。別の一例では、選択された参照モデル 9 2 3 である「混合分布数 1 6 の成人男性女
20 モデル」を統計量の初期値として統計量記憶部 9 0 6 c に記憶する。具体的には、初期標準モデル作成部 9 0 6 b は、上記式 1 3 に示される出力分布を生成する。

そして、統計量推定部 9 0 6 d は、参照モデル選択部 9 0 5 が選択した 2 個の参照モデル 9 2 3 を用いて、統計量記憶部 9 0 6 c に記憶された
25 た標準モデルの統計量を推定する（図 4 のステップ S 1 0 2 c）。つまり、2 個（ $Ng = 2$ ）の参照モデル 9 2 3 における出力分布、即ち、上記式

19に示される出力分布に対する標準モデルの確率（ここでは、上記式
25に示される尤度 $\log P$ ）を極大化もしくは最大化するような標準モ
デルの統計量（上記式16に示される混合重み係数、上記式17に示さ
れる平均値、及び、上記式18に示される分散値）を推定する。ただし、
5 本実施の形態では、上記式19に示された出力分布における式21は、
16（各参照モデルの混合分布数）である。

具体的には、上記式26、式27及び式28に従って、それぞれ、標
準モデルの混合重み係数、平均値及び分散値を算出する。

このとき、統計量推定部906dの第3近似部906eは、標準モデ
10 ルの各ガウス分布はお互いに影響を与えないと仮定して、式53の近似
式を用いる。また、繰り返し回数Rが1回目の場合には、式54に示さ
れる標準モデルのガウス分布の近傍の式55とは、式54が示す出力分
布とのマハラノビス距離、カルバック・ライブラー（KL）距離などの
分布間距離が最も近いものと2番目に近いものの2個（近傍指示パラメ
15 ータ $G=2$ ）の式56に示される参照モデル923のガウス分布が存在
する空間であると近似する。一方、繰り返し回数Rが2回目以上の場合
には、式54に示される標準モデルのガウス分布の近傍の式55とは、
式54が示す出力分布とのマハラノビス距離、カルバック・ライブラー
（KL）距離などの分布間距離が最も近いものの1個（近傍指示パラメ
20 ータ $G=1$ ）の式56に示される参照モデル923のガウス分布が存在す
る空間であると近似する。

以上の第3近似部906eによる近似式を考慮してまとめると、統計
量推定部906dでの計算式は、次の通りになる。つまり、統計量推定
部906dは、式59、式60及び式61に従って、それぞれ、混合重
25 み係数、平均値及び分散値を算出し、それらのパラメータによって特定
される標準モデルを最終的な標準モデル922として生成する。ただし、

第 3 の実施の形態における第 2 の方法である、混合重み係数の値をゼロにして、平均値をゼロ、分散値を 1 にする方法を用いる。また、繰り返し回数に対応して近傍指示パラメータ G の値は異なる。なお、近傍指示パラメータ G の値に依存して、上記の方法を、第 3 の実施の形態における第 1 から第 3 の方法のいずれかに決定してもよい。

統計量推定部 906d は、このように推定した標準モデルの統計量を統計量記憶部 906c に記憶する。そして、このような統計量の推定と統計量記憶部 906c への記憶を R (≥ 1) 回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル 922 の統計量として出力する。

図 38 に、第 3 近似部 906e を用いて作成した標準モデル 922 を用いた認識実験の結果を示す。縦軸に成人 (男性と女性) の認識率 (%)、横軸に繰り返し回数 R を示す。繰り返し回数 $R = 0$ とは、学習を行う前での初期標準モデル作成部 906b が作成した初期モデルにより認識した結果である。また、繰り返し回数 $R = 1$ のときは、近傍指示パラメータ $G = 2$ とし、繰り返し回数 $R = 2 \sim 5$ のときは、近傍指示パラメータ $G = 1$ とした。

グラフ「データ」は、数日間かけて音声データより学習した場合の結果を表しており、グラフ「女性」、グラフ「男性」は、それぞれ、初期モデルを成人女性、成人男性としたときの結果を表している。参照モデルによる本発明による学習時間は数十秒のオーダーであった。実験結果より、短時間に高い精度の標準モデルが作成できていることがわかる。

ここで、参考のために、図 39 に、第 3 の実施の形態における第 2 近似部 306e により作成された標準モデルによる認識率を示す。本実施の形態における第 3 近似部 906e と異なるのは、繰り返し回数 R によらず近傍指示パラメータ $G = 1$ であるということである。実験結果より、

初期モデルとして成人女性を選択すると良好な結果が得られることがわかる。また、初期モデルとして成人男性を選択すると、精度が少し劣化していることがわかる。図 38 の結果とあわせると、第 3 近似部 906 e による標準モデルは初期モデルに依存せずに高い精度の標準モデルが
5 作成できていることがわかる。

以上説明したように、本発明の第 8 の実施の形態によれば、類似度情報に基づいて参照モデルを準備するため、利用情報及び仕様情報にふさわしい参照モデルを必要なタイミングで準備することができる。また、近傍指示パラメータ G を繰り返し回数 R によって変化させることで、初
10 期モデルにかかわらず精度の高い標準モデルを提供することができる。

なお、統計量推定部 906 d による処理の繰り返し回数は、上記式 25 に示された尤度の大きさがある一定のしきい値以上になるまでの回数としてもよい。

また、標準モデル 922 は、音素ごとに HMM を構成するに限らず、
15 文脈依存の HMM で構成してもよい。

また、標準モデル作成部 906 は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル 922 を構成する HMM は、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布
20 により構成してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などの HMM の構造や、状態数などを決定してもよい。

全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がPDA (Personal Digital Assistant) 1001に組み込まれた例が示されている。以下、本実施の形態では音声認識用の標準モデルを作成する場合を例にして説明する。

5 PDA 1001は、携帯情報端末であり、事象の集合と事象又は事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、参照モデル記憶部1003と、標準モデル作成部1006と、アプリ・仕様情報対応データベース1014と、マイク1012と、音声認識部1013とを備える。標準モデル作成部1006は、標準モデル構造決定部1006aと、初期標準モデル作成部1006bと、統計量記憶部306cと、統計量推定部306dとを備える。

標準モデル作成部1006は、送信されたアプリ起動情報1027(ここでは、起動したアプリケーションのID番号)に基づいて、アプリ・仕様情報対応データベース1014を用いて、仕様情報1025を取得する。図41は、仕様情報対応データベース1014のデータ例を示す。仕様情報対応データベース1014には、アプリケーション(ID番号及び名前)に対応する仕様情報(ここでは、混合分布数)が登録されている。

20 標準モデル作成部1006は、取得した仕様情報1025に基づいて、参照モデル記憶部1003が記憶した1個の参照モデル1021に対する確率又は尤度を最大化又は極大化するように標準モデル1022を作成する処理部であり、第3の実施の形態における第2近似部306eの機能を有する。

25 音声認識部1013は、標準モデル作成部1006で作成された標準

を認識する。

次に、以上のように構成されたPDA1001の動作について説明する。

図42は、PDA1001の動作手順を示すフローチャートである。

5 ここで、参照モデル記憶部1003には、あらかじめ多くの混合分布数をもつ利用者用モデルが参照モデル1021として1個、記憶されているとする。参照モデル1021は、音素ごとのHMMにより構成される。参照モデル1021の一例を図43に示す。この参照モデルは、状態数3個、各状態は分布数が300個の混合ガウス分布によりHMMの
10 出力分布が構成される。特徴量として、12次元のメルケプストラム係数、12次元のデルタメルケプストラム係数、デルタパワーの合計25次元（J=25）の特徴量が用いられる。

まず、利用者は、例えば「株取引」というアプリケーションを起動する（ステップS1000）。

15 これに対して、標準モデル作成部1006は、アプリ起動情報として起動されたアプリケーションのID「3」を受信する（ステップS1001）。そして、アプリ・仕様情報対応データベース1014を用いてID「3」に対応する仕様情報1025である「混合分布数126」に基づいて、標準モデル1022を作成する（ステップS1002）。具体的
20 には、標準モデル1022として、混合分布数126（Mf=126）で、3状態の文脈依存型のHMMにより構成する。

次に、標準モデル作成部1006は、仕様情報1025を受信して（ステップS1001）、仕様情報1025に基づいて標準モデルを作成する（ステップS1002）。

25 最後に、音声認識部1013は、標準モデル作成部1006によって作成された標準モデル1022に従って、マイク1012から入力され

た利用者の音声を認識する（ステップS1003）。

次に、図42におけるステップS1002（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図4に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等
5 が異なる。

まず、標準モデル構造決定部1006aは、アプリ起動情報1027としてアプリケーションID「3」を受信した後に、アプリ・仕様情報対応データベース1014を用いてID「3」に対応した仕様情報1025（「混合分布数126」）を参照することにより、標準モデルの構造
10 を混合分布数126（ $Mf=126$ ）で、3状態の文脈依存型のHMMと決定する（図4のステップS102a）。

そして、初期標準モデル作成部1006bは、標準モデル構造決定部1006aが決定した標準モデルの構造に基づいて、標準モデルを計算するための統計量の初期値を決定する（図4のステップS102b）
15 こでは、k-means法とマハラノビス汎距離を用いた方法により、後述するクラスタリングを行ったものを統計量の初期値として統計量記憶部306cに記憶する。

そして、統計量推定部306dは、参照モデル記憶部1003に格納された参照モデル1021を用いて、統計量記憶部306cに記憶された標準モデルの統計量を推定する（図4のステップS102c）。なお、
20 この統計量推定部306dによる推定処理は、第3の実施の形態と同様である。

次に、初期標準モデル作成部1006bによる初期値の決定方法、つまり、k-means法とマハラノビス汎距離を用いた方法によるクラスタリングについて説明する。図44にクラスタリングのフローチャートを示
25 す。また、図45～図48にクラスタリングのイメージ図を示す。

まず、図４４のステップＳ１００４において、標準モデルの混合分布数である１２６個の代表点を準備する（図４５）。ここでは、参照モデルの３００個の出力分布の中から１２６個の出力分布を選択して、選択された分布の平均値を代表点とする。

５ 次に、図４４のステップＳ１００５において、各代表点にマハラノビス汎距離が近い参照モデルの出力ベクトルを決定する（図４６）。そして、図４４のステップＳ１００６において、ステップＳ１００５で決定した近い分布を１つのガウス分布で表現して平均値を新しい代表点とする（図４７）。

１０ 続いて、図４４のステップＳ１００７において、クラスタリング操作を停止するかどうかを決定する。ここでは、各代表点と参照ベクトルの分布とのマハラノビス汎距離の変化率（１回前の代表点との距離との差分）がしきい値以下になった場合に停止とする。停止条件を満たさない場合、図４４のステップＳ１００５に戻り、近い分布を決定して同様の
１５ 操作を繰り返す。

一方、停止条件を満たす場合には、図４４のステップＳ１００８に進み、統計量の初期値を決定して統計量記憶部３０６ｃに記憶する。このようにして、クラスタリングによる初期値の決定が行われる。

以上説明したように、本発明の第９の実施の形態によれば、アプリケーションに連動して自動的に仕様情報にふさわしい標準モデルを獲得
２０ することができる。

なお、標準モデル１０２２は、音素ごとにＨＭＭを構成してもよい。

また、標準モデル作成部１００６は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

２５ また、標準モデル１０２２を構成するＨＭＭは、音素ごとに異なる状態数により構成してもよい。音素ごとに異なる分布数の混合ガウス分

布により構成してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

(第10の実施の形態)

図49は、本発明の第10の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ801に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル(適応モデル)を作成する場合を例にして説明する。

サーバ801は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部711と、参照モデル準備部702と、参照モデル記憶部703と、利用情報受信部704と、参照モデル選択部705と、標準モデル作成部706と、仕様情報受信部707と、標準モデル記憶部708と、標準モデル送信部709と、参照モデル受信部810とを備える。

参照モデル準備部702は、読み込み部711が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部703へ送信する。参照モデル記憶部703は、送信された参照モデル721を記憶する。また、参照モデル準備部702は、端末装置712からの送信に対して参照モデル受信部810が受信した音声認識用参照モデルを参照モデル記憶部703へ送信する。参照モデル記憶部703は、送信された参照モデル

仕様情報受信部 707 は、端末装置 712 から仕様情報 725 を受信する。利用情報受信部 704 は、端末装置 712 から利用情報 724 である雑音下で発声した利用者の音声を受信する。参照モデル選択部 705 は、利用情報受信部 704 が受信した利用情報 724 である利用者の
5 音声に音響的に近い話者・雑音・声調子の参照モデル 723 を、参照モデル記憶部 703 が記憶している参照モデル 721 から選択する。

標準モデル作成部 706 は、仕様情報 725 に基づいて、参照モデル選択部 705 が選択した参照モデル 723 に対する確率又は尤度を最大化又は極大化するように標準モデル 722 を作成する処理部であり、第
10 2 の実施の形態における標準モデル作成部 206 と同一の機能を有する。標準モデル記憶部 708 は、仕様情報 725 に基づいた 1 もしくは複数の標準モデルを記憶する。標準モデル送信部 709 は、利用者の端末装置 712 から、仕様情報 725 と標準モデルの要求信号とを受信すると、その端末装置 712 へ、仕様に適した標準モデルを送信する。

15 次に、以上のように構成されたサーバ 801 の動作について説明する。

図 50 は、サーバ 801 の動作手順を示すフローチャートである。なお、このサーバ 801 の動作手順を説明するための参照モデル及び標準モデルの一例は、第 7 に実施の形態における図 31 と同様である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備
20 する（図 50 のステップ S800、S801）。つまり、参照モデル準備部 702 は、読み込み部 711 が読み込んだ、CD-ROM などのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部 703 へ送信し、参照モデル記憶部 703 は、送信された参照モデル 721 を記憶する（図 50 のステップ S80
25 0）。ここでは、参照モデル 721 は、話者・雑音・声の調子ごとに、音

端末装置 712 が送信して参照モデル受信部 810 が受信した、利用者と端末装置 712 に適した音声認識用参照モデルを参照モデル記憶部 703 へ送信し、参照モデル記憶部 703 は、送信された参照モデル 721 を記憶する(図 50 のステップ S801)。ここでは、各参照モデルは、
5 図 31 の参照モデル 721 に示されるように、状態数 3 個、各状態は混合分布数が 128 個の混合ガウス分布により HMM の出力分布が構成される。特徴量として 25 次元 ($J=25$) のメルケプストラム係数が用いられる。

以下、これらの参照モデル 721 を用いた標準モデル 722 の作成及び端末装置 712 への送信(図 50 のステップ S802 ~ S809)は、
10 第 7 の実施の形態における手順(図 30 のステップ S701 ~ S708)と同様である。

このようにして、端末装置 712 に記憶された自分用モデルをサーバにアップロードして標準モデル作成の材料にすることができるので、例えば、サーバ 801 において、アップロードされてきた参照モデルと既に保持している他の参照モデルとを統合して更に混合数の多い高精度の標準モデルを作成し、端末装置 712 にダウンロードして利用することが可能となる。したがって、端末装置 712 に簡易的な適応機能が付属され、簡易的に適応したモデルをアップロードして、さらに高精度な標準モデルを作成することもできる。
20

図 51 は、本実施の形態における標準モデル作成装置を具体的に適用したシステム例を示す図である。ここには、インターネットや無線通信等を介して通信し合うサーバ 701 と端末装置 712 (携帯電話機 712a、カーナビゲーション装置 712b) とが示されている。

25 たとえば、携帯電話機 712a は、利用者の音声を利用情報とし、携

とし、予め記憶しているサンプルモデルを参照モデルとし、それら利用
情報、仕様情報及び参照モデルをサーバ701に送信することで、標準
モデルの作成を要求する。その要求に対してサーバ701で標準モデル
が作成されると、携帯電話機712aは、その標準モデルをダウンロー
5 ドし、その標準モデルを用いて利用者の音声を認識する。例えば、利用
者の音声、内部に保持するアドレス帳の名前と一致した場合には、そ
の名前に対応する電話番号に自動発呼する。

また、カーナビゲーション装置712bは、利用者の音声を利用情報
とし、カーナビゲーション装置での利用である旨（CPUの処理能力が
10 通常であること）を仕様情報とし、予め記憶しているサンプルモデルを
参照モデルとし、それら利用情報、仕様情報及び参照モデルをサーバ7
01に送信することで、標準モデルの作成を要求する。その要求に対し
てサーバ701で標準モデルが作成されると、カーナビゲーション装置
712bは、その標準モデルをダウンロードし、その標準モデルを用い
15 て利用者の音声を認識する。例えば、利用者の音声、内部に保持する
地名と一致した場合には、その地名を目標点とする現地点からの道順を
示す地図を画面に自動表示する。

このようにして、携帯電話機712a及びカーナビゲーション装置7
12bは、自装置に適した標準モデルの作成をサーバ701に依頼する
20 ことで、標準モデルの作成に必要な回路や処理プログラムを自装置内に
実装する必要がなくなるとともに、様々な認識対象の標準モデルを必要
なタイミングで獲得することができる。

以上説明したように、本発明の第10の実施の形態によれば、参照モ
デル受信部810が受信した参照モデルを利用して標準モデルを作成で
25 きるため、精度の高い標準モデルが提供される。つまり、端末装置71

側で保持する参照モデルのバリエーションが増加し、他の人が利用したときにさらに高精度の標準モデルを提供することができる。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

- 5 なお、参照モデル受信部 810 は、端末装置 712 とは異なる他の端末装置から参照モデルを受信してもよい。

また、図 51 に示された応用例は、本実施の形態に限られるものではなく、他の実施の形態にも適用することができる。つまり、第 1 ～ 第 9 の実施の形態で作成された標準モデルを各種記録媒体や通信を介して
10 様々な電子機器に配信することで、それらの電子機器において、制度の高い音声認識、画像認識、意図理解等を行うことが可能となる。さらに、上記実施の形態における標準モデル作成装置を各種電子機器に内蔵させることで、音声認識、画像認識、意図理解等の認識・認証機能を備えるスタンドアローンの電子機器を実現することもできる。

- 15 以上、本発明に係る標準モデル作成装置について、実施の形態に基づいて説明したが、本発明は、これらの実施の形態に限定されるものではない。

たとえば、第 1 ～ 第 10 の実施の形態における標準モデルの統計量の近似計算については、各実施の形態における近似計算だけに限られず、
20 第 1 ～ 第 4 の実施の形態における合計 4 種類の近似計算の少なくとも 1 つを用いてもよい。つまり、4 種類の近似計算のいずれであってもよいし、2 以上の種類の近似計算の組み合わせであってもよい。

また、第 2 の実施の形態では、統計量推定部 206d の一般近似部 206e は、標準モデルの混合重み係数、平均値及び分散値を、それぞれ、
25 式 45、式 46 及び式 47 に示される近似式に従って算出したが、これ

式を用いて算出してもよい。

(式 6 3)

$$\omega_{f(m)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, k) \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

($m = 1, 2, \dots, M_f$)

(式 6 4)

$$\mu_{f(m,j)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} x_{(j)} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

($m = 1, 2, \dots, M_f, j = 1, 2, \dots, J$)

5

(式 6 5)

$$\sigma_{f(m,j)}^2 \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} (x_{(j)} - \mu_{f(m,j)})^2 \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \nu_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

($m = 1, 2, \dots, M_f, j = 1, 2, \dots, J$)

このような近似式を用いて作成した標準モデルによれば、高い認識性能が得られることが発明者らによって確認されている。たとえば、参照

適応前では 82.2%であったものが、上記非特許文献 2 に示された十分統計量による方法では、85.0%、上記近似式による方法では 85.5% に改善された。つまり、十分統計量による方法と比べ、高い認識性能が獲得できていることがわかる。また、参照モデルの混合数を 64、標準モデルの混合数を 16 とした場合についての認識結果は、上記近似式による方法では、85.7% と高い認識率が獲得できている。

また、初期標準モデル作成部による初期標準モデルの作成においては、図 5 2 に示されるようなクラス ID・初期標準モデル・参照モデル対応表を予め準備しておき、この表に従って、初期標準モデルを決定してもよい。以下、このようなクラス ID・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定方法について説明する。なお、クラス ID とは、標準モデルを用いた認識対象の種別を識別する ID であり、標準モデルの種類に対応する。

図 5 2 に示されたクラス ID・初期標準モデル・参照モデル対応表は、一定の共通する性質を有する複数の参照モデルに対して、それらを識別する 1 つのクラス ID を対応づけるとともに、それら参照モデルと共通する性質を持つ予め作成された初期標準モデルを対応づけた表である。この表では、参照モデル 8 A A ~ 8 A Z に対して、クラス ID 及び初期標準モデル 8 A が対応づけられ、参照モデル 6 4 Z A ~ Z Z に対して、クラス ID 及び初期標準モデル 6 4 Z が対応づけられている。標準モデル作成部は、使用する参照モデルの性質と共通する初期標準モデルを使用することによって、精度の高い標準モデルを生成することができる。

ここで、クラス ID、初期標準モデル及び参照モデルの添え字記号 8 A、8 A A における最初の記号「8」等は、混合分布数を意味し、2 番目の記号「A」等は、大分類、例えば、騒音下における音声認識の場合であれば、騒音環境の種類（家庭内騒音下を A、電車内騒音下を B など）

を意味し、3番目の記号「A」等は小分類、例えば、音声認識の対象となる人の属性（低学年の小学生をA、高学年の小学生をBなど）を意味する。したがって、図52のクラスID・初期標準モデル・参照モデル対応表における参照モデル8AA～AZは、図53に示されるような混合分布数8のモデルであり、参照モデル64ZA～ZZは、図54に示されるような混合分布数64のモデルであり、初期標準モデル8A～64Zは、図55に示されるような混合分布数8～16のモデルである。

次に、このようなクラスID・初期標準モデル・参照モデル対応表の作成方法を説明する。図56は、その手順を示すフローチャートであり、図57～図60は、各ステップでの具体例を示す図である。ここでは、騒音環境下での音声認識を例とし、表だけでなく、クラスID、初期標準モデル及び参照モデルも含めて新規に作成する場合の手順を説明する。

まず、音声データを音響的に近いグループに分類する（図56のステップS1100）。たとえば、図57に示されるように、音声データを利用情報である雑音環境で分類する。環境A（家庭内騒音下での音声データ）には、家庭内騒音下で収録した小学生低学年の音声、小学生高学年の音声、成人女性の音声などが含まれ、環境B（電車内での音声データ）には、電車内で収録した小学生低学年の音声、小学生高学年の音声、成人女性の音声などが含まれるように分類する。なお、利用情報である話者の性別、年齢層、笑い声・怒った声などの声の性質、読み上げ調・会話調などの声の調子、英語・中国語などの言語などで分類してもよい。

次に、仕様情報等に基づいて、準備する参照モデルの1以上のモデル構造を決定する（図56のステップS1101）。たとえば、8混合、16混合、32混合及び64混合を対象とすることを決定する。なお、モデル構造の決定においては、混合分布数を決定するに限らず、HMMの

もよい。

続いて、初期標準モデルを作成する（図５６のステップＳ１１０２）。つまり、上記音声データの分類（ステップＳ１１００）において決定した分類（環境Ａ、環境Ｂ、…）ごとに、ステップＳ１１０１において決定したモデル構造ごとの初期標準モデルを作成する。例えば、図５８に示されるように、初期標準モデル８Ａであれば、８混合の初期標準モデルを、家庭内騒音下（環境Ａ）における音声データ（低学年の小学生、高学年の小学生、成人男、成人女等の音声データ）を用いて、バウム・ウェルチアルゴリズムなどにより学習して作成する。

次に、参照モデルを作成する（図５６のステップＳ１１０３）。つまり、上記ステップＳ１１０２において作成した初期標準モデルを用いて参照モデルを作成する。具体的には、参照モデルを学習する音声データの雑音環境と同じ雑音環境で学習した、同じ混合分布数をもつ初期標準モデルを用いて参照モデルを学習する。例えば、図５９に示されるように、参照モデル８ＡＡは、混合分布数８の家庭内騒音下での小学生低学年の音声データで学習するモデルであり、学習を行う際の初期値として、同じ環境である家庭内騒音下での音声データ（小学生低学年、小学生高学年、成人女性、成人男性の音声を含む）で学習した初期標準モデルを用いる。学習方法として、バウム・ウェルチアルゴリズムを用いる。

最後に、クラスＩＤを付与する（図５６のステップＳ１１０４）。たとえば、騒音環境下ごとに１つのクラスＩＤを付与することによって、図６０に示されるクラスＩＤ・初期標準モデル・参照モデル対応表、つまり、"クラスＩＤ付き初期標準モデル"及び"クラスＩＤ付き参照モデル"が作成される。

なお、このようなクラスＩＤ・初期標準モデル・参照モデル対応表は、作成された表として予め端末（標準モデル作成装置）が保持している。

要はない。端末（標準モデル作成装置）は、図 6 1 に示されるように、他の装置（サーバ）と通信することによって表を完成させてもよい。つまり、標準モデル作成装置（端末）は、通信網などを介して、"クラス ID 付き初期標準モデル"、"クラス ID 付き参照モデル"を取得することが可能である。もともと、端末は必ずしも"クラス ID 付き初期標準モデル"、"クラス ID 付き参照モデル"を取得する必要はなく事前に記憶させて出荷してもよい。

図 6 1 に示されるように、端末は、以下のような方法によって、"クラス ID 付き初期標準モデル"、"クラス ID 付き参照モデル"を取得することができる。第 1 の方法として、端末は、"クラス ID 付き初期標準モデル"（例えば規格化コンソーシアムなどで事前に定義されたクラス ID のつけ方に遵守したもの）を記憶しているケースである。このとき、端末は、1 以上のサーバから"クラス ID 付き参照モデル"（例えば規格化コンソーシアムなどで事前に定義されたクラス ID のつけ方に遵守したもの）をダウンロードする。なお、端末に、"クラス ID 付き参照モデル"を出荷時に記憶させておいてもよい。

また、第 2 の方法として、端末は、"クラス ID 付き初期標準モデル"を記憶していないケースである。このとき、端末は、サーバ（図 6 1 のサーバ 1）から"クラス ID 付き初期標準モデル"をダウンロードする。次に、端末は、1 以上のサーバ（図 6 1 のサーバ 2）から"クラス ID 付き参照モデル"をダウンロードする。必要に応じて逐次的にクラス ID の定義の追加、変更が可能である。また、端末のメモリの節約にもなる。

さらに、第 3 の方法として、端末は、クラス ID と初期標準モデル・参照モデルの対応関係を明記した"クラス ID ・初期標準モデル・参照モデル対応表"を記憶しているケースである。このとき、端末は、"対応表"記憶していないサーバ（図 6 1 のサーバ 3）に"対応表"をアップロード

する。サーバは、送信された"対応表"に基づき"クラスID付き参照モデル"を準備する。端末は、準備された"クラスID付き参照モデル"をダウンロードする。

次に、このようなクラスID・初期標準モデル・参照モデル対応表を用いた初期標準モデル作成部による初期標準モデルの決定方法について説明する。図62は、その手順を示すフローチャートである。図63及び図64は、各ステップでの具体例を示す図である。

まず、標準モデルの作成に用いる参照モデルからクラスIDを抽出する(図62のステップS1105)。たとえば、図63に示されるテーブルに従って、選択された参照モデルから、対応するクラスIDを抽出する。ここでは、抽出したクラスIDとして、8Aが1個、16Aが3個、16Bが1個、64Bが1個とする。

次に、抽出したクラスIDを用いて標準モデル作成に用いる初期標準モデルを決定する(図62のステップS1106)。具体的には、以下の手順に従って初期標準モデルを決定する。

(1) 作成する標準モデルの混合分布数(16混合)と同じクラスID(16*)をもつ参照モデルから抽出したクラスID(16A、16B)に着目し、その中から一番多く抽出されたクラスIDに対応する初期標準モデルを最終的な初期標準モデルと決定する。たとえば、標準モデルの構造が16混合の場合には、16混合に関するクラスIDとして、16Aが3個、16Bが1個抽出されているので、クラスIDが16Aの初期標準モデルを採用する。

(2) 作成する標準モデルの混合分布数(8混合)と同じクラスID(8*)をもつ参照モデルから抽出したクラスID(8A)に着目し、同じクラスIDをもつ初期標準モデルを最終的な初期標準モデルと決定する。

スIDとして、8Aが1個抽出されているので、クラスIDが8Aの初期標準モデルを採用する。

(3) 作成する標準モデルの混合分布数(32混合)と同じクラスID(32*)をもつ参照モデルから抽出したクラスIDに着目し、存在しない場合、仕様情報に着目してその中から一番多く抽出されたクラスID(*A)をもつ初期標準モデル(8A、16A)を用いてクラスタリングにより32混合にして最終的な初期標準モデルとする(図44を参照)。たとえば、標準モデルの構造が32混合の場合には、32混合に関するクラスIDが抽出されていないので、一番多く抽出されたクラスID(16A)を用いてクラスタリングにより32混合にして初期標準モデルとする。

なお、はじめに作成する標準モデルの仕様情報(混合分布数など)に着目せず、利用情報(雑音の種類など)に着目して初期値を決定してもよい。

図64に、第3近似部を用いて作成した混合分布数が64の標準モデルを用いた認識実験の結果を示す。縦軸に成人(男性と女性)の認識率(%),横軸に繰り返し回数Rを示す。繰り返し回数R=0とは、学習を行う前での初期標準モデル作成部が作成した初期モデルにより認識した結果である。また、繰り返し回数R=1~5において、近傍指示パラメータG=1とした。

グラフ「データ」は、数日間かけて音声データより学習した場合の結果を表しており、グラフ「女性」、グラフ「男性」は、それぞれ、初期モデルを成人女性、成人男性としたときの結果を表している。参照モデルによる本発明による学習時間は数分のオーダーであった。この実験結果より、成人女性の参照モデルを初期標準モデルと決定した場合には、音声データで学習した結果よりも高い精度の標準モデルが作成できている

ことが分かる。

このことは、音声データを分割し、分割した音声データをそれぞれの参照モデルとして厳密に学習したのちに統合したほうが、音声データによる学習の課題である局所解に陥るという問題を解決できる可能性を示している（音声データによる学習との認識精度での比較）。

また、音声データの収録が困難な子供の音声データに対しては、データ数に適切である混合分布数の少ない参照モデルで厳密に学習して、多くの音声データの収録が可能な成人の音声データに対しては、混合分布数の多い参照モデルで厳密に学習して、そのあとで本発明により統合して標準モデルを作成すれば、極めて精度の高い標準モデルが作成できることが期待できる。

なお、標準モデルの混合分布数が16の場合における認識実験（図39）では、本発明による方法は、音声データで学習した標準モデルの認識率を超えていない。このことは、音声データを16混合の参照モデルの形にしたときに音声データの情報が欠如したためだと考えられる。参照モデルを64混合で作成して音声データの特徴を十分保持しておけばより高い精度の標準モデルが作成できる。このことより、第9の実施の形態では、参照モデルの混合分布数を300と大きめに設定している。

また、図39及び図64に示される認識実験より、初期標準モデルが認識精度に与える影響が示されており、初期標準モデルの決定方法の重要性を物語っている（図64において、成人女性の参照モデルを初期標準モデルとして利用した場合、成人男性の参照モデルを利用する場合より高い精度の標準モデルが作成できることが示されている）。

以上のように、クラスID・初期標準モデル・参照モデル対応表に従って、参照モデルと共通する性質の初期標準モデルを用いることで、精度の高い標準モデルを作成することができる。

なお、このようなクラスID・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定は、上記実施の形態1～10のいずれにおいても採用することができる。

また、上記実施の形態では、標準モデルの統計量を推定する際に、参照モデルに対する標準モデルの尤度として式25が用いられたが、本発明はこのような尤度関数に限られず、例えば、以下の式66に示される尤度関数を用いてもよい。

(式66)

$$\log L = \sum_{i=1}^N \int_{-\infty}^{\infty} \log \left[\sum_{m=1}^M \omega_{(m)} f(x, \mu_{(m)}, \sigma_{(m)}^2) \right] \alpha_{(i)} \left\{ \sum_{l=1}^{L_i} \nu_{(l)} g_i(x, \mu_{(l)}, \sigma_{(l)}^2) \right\} dx$$

ここで、 $\alpha(i)$ は、統合する各参照モデル*i*に対応した重要度を示す重み付けである。たとえば、音声認識における話者適用であれば、重要度は、利用者の音声と統合モデルを作成した音声の近さにより決定される。つまり、参照モデルが利用者の音声に近い（重要度が大きい）場合に、 $\alpha(i)$ は大きな値に設定される（大きく重み付けされる）。統合モデルと利用者の音声との近さは、利用者の音声を統合モデルに入力したときの尤度の大きさにより決定すればよい。これによって、複数の参照モデルを統合して標準モデルを作成する際に、利用者の音声に近い参照モデルほど大きな重み付けで標準モデルの統計量に影響を与えることとなり、より利用者の特性を反映した精度の高い標準モデルが作成される。

また、各実施の形態における標準モデル構造決定部は、利用情報や仕様情報などの各種要因に基づいて標準モデルの構造を決定したが、本発明は、これらの要因だけに限られず、例えば、音声認識の場合であれば、認識の対象となる人の年齢、性別、声質の話者性、感情又は健康状態に基づく声の調子、発話速度、発話の丁寧さ、方言、背景雑音の種類、背景雑音の大きさ、音声と背景雑音とのSN比、マイク特性及び認識語彙

の複雑さなどの各種属性に依存して標準モデルの構造を決定してもよい。

具体的には、図 6 5 (a) ~ (j) に示されるように、音声認識の対象となる人の年齢が高いほど標準モデルを構成するガウス分布の数（混合数）を大きくしたり（図 6 5 (a)）、音声認識の対象となる人が男性の場合には女性の場合よりも大きな混合数にしたり（図 6 5 (b)）、音声認識の対象となる人の音質が「通常」よりも「ハスキー」、さらに「しわがれ声」となるほど混合数を大きくしたり（図 6 5 (c)）、音声認識の対象となる声の感情による調子が「通常」よりも「怒り声」、さらに「泣き／笑いながらの声」となるほど混合数を大きくしたり（図 6 5 (d)）、音声認識の対象となる人の発話速度が速く／遅くなるほど混合数を大きくしたり（図 6 5 (e)）、音声認識の対象となる人の発話の丁寧さが「朗読調」よりも「講演調」、さらに「会話調」となるほど混合数を大きくしたり（図 6 5 (f)）、音声認識の対象となる人の方言が「標準語」よりも「大阪弁」、さらに「鹿児島弁」となるほど混合数を大きくしたり（図 6 5 (g)）、音声認識における背景雑音が大きくなるほど混合数を小さくしたり（図 6 5 (h)）、音声認識に使用するマイクの性能が高くなるほど混合数を大きくしたり（図 6 5 (i)）、音声認識の対象となる語彙が増加するほど混合数を大きくしたり（図 6 5 (j)）すればよい。これらの例の多くは、認識対象の音声のばらつきが大きいほど、混合数を大きくして精度を確保するという観点から混合数が決定される。

産業上の利用の可能性

本発明に係る標準モデル作成装置は、確率モデル等を用いた音声、文字、画像等の対象物を認識する装置等として利用することができ、例えば、音声によって各種処理を実行するテレビ受信装置・カーナビゲーション装置、音声を他の言語に翻訳する翻訳装置、音声で操作するが、

装置、音声による検索キーワードで情報を検索する検索装置、人物検出・指紋認証・顔認証・虹彩認証等を行う認証装置、株価予測、天気予測などの予測を行う情報処理装置等として利用することができる。

請 求 の 範 囲

1. 音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、
- 5 一定の属性を有する音声の特徴を示す確率モデルである 1 以上の参照モデルを記憶する参照モデル記憶手段と、
- 前記参照モデル記憶手段に格納された 1 以上の参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備え、
- 10 前記標準モデル作成手段は、
- 作成する標準モデルの構造を決定する標準モデル構造決定部と、
- 構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成部と、
- 初期値が決定された標準モデルの前記参照モデルに対する確率又は尤
- 15 度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有する
- ことを特徴とする標準モデル作成装置。
2. 前記標準モデル作成装置はさらに、
- 20 音声認識の対象となる属性に関する情報である利用情報に基づいて、
- 前記参照モデル記憶手段に記憶されている参照モデルの中から 1 以上の参照モデルを選択する参照モデル選択手段を備え、
- 前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルの統計量を用いて標準モデルを作成する
- 25 ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

3. 前記標準モデル作成装置はさらに、

前記利用情報を作成する利用情報作成手段を備え、

前記参照モデル選択手段は、作成された利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から 1 以上の参照モ

5 デルを選択する

ことを特徴とする請求の範囲 2 記載の標準モデル作成装置。

4. 前記標準モデル作成装置には通信路を介して端末装置が接続され、

前記標準モデル作成装置はさらに、

10 前記端末装置から前記利用情報を受信する利用情報受信手段を備え、

前記参照モデル選択手段は、受信された利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から 1 以上の参照モデルを選択する

ことを特徴とする請求の範囲 2 記載の標準モデル作成装置。

15

5. 前記標準モデル構造決定部は、作成する標準モデルの仕様に関する情報である仕様情報、及び、音声認識の対象となる属性に関する情報である利用情報の少なくとも一方に基づいて、前記標準モデルの構造を決定する

20 ことを特徴とする請求の範囲 1 記載の音声認識用の標準モデル作成装置。

6. 前記仕様情報とは、標準モデルを使用するアプリケーションプログラムの種類、及び、標準モデルを使用する機器の仕様の少なくとも一方

25 の仕様を示す

ことを特徴とする請求の範囲 5 記載の音声認識用の標準モデル作成装

置。

7. 前記属性とは、年齢、性別、声質の話者性、感情又は健康状態に基づく声の調子、発話速度、発話の丁寧さ、方言、背景雑音の種類、背景
- 5 雑音の大きさ、音声と背景雑音との SN 比、マイク特性及び認識語彙の複雑さの少なくとも 1 つに関する情報を含む

ことを特徴とする請求の範囲 5 記載の音声認識用の標準モデル作成装置。

- 10 8. 前記標準モデル作成装置はさらに、

標準モデルを使用するアプリケーションプログラムと標準モデルの仕様との対応を示すアプリケーション仕様対応データベースを前記仕様情報として保持する仕様情報保持手段を備え、

- 前記標準モデル構造決定部は、前記仕様情報保持手段に保持されたアプリケーション仕様対応データベースから、起動されるアプリケーション
- 15 プリケーション仕様対応データベースから、起動されるアプリケーションプログラムに対応する仕様を読み出し、読み出した仕様に基づいて、前記標準モデルの構造を決定する

ことを特徴とする請求の範囲 5 記載の標準モデル作成装置。

- 20 9. 前記標準モデル作成装置はさらに、

前記仕様情報を作成する仕様情報作成手段を備え、

前記標準モデル構造決定部は、作成された仕様情報に基づいて、前記標準モデルの構造を決定する

ことを特徴とする請求の範囲 5 記載の標準モデル作成装置。

25

10. 前記標準モデル作成装置には通信路を介して端末装置が接続され、

前記標準モデル作成装置はさらに、

前記端末装置から前記仕様情報を受信する仕様情報受信手段を備え、

前記標準モデル構造決定部は、受信された仕様情報に基づいて、前記標準モデルの構造を決定する

5 ことを特徴とする請求の範囲 5 記載の標準モデル作成装置。

1 1. 前記参照モデル及び前記標準モデルは、1 以上のガウス分布を用いて表現され、

前記標準モデル構造決定部は、前記標準モデルの構造として、少なくともガウス分布の混合数を決定する

10

ことを特徴とする請求の範囲 5 記載の標準モデル作成装置。

1 2. 前記標準モデル作成装置には、通信路を介して端末装置が接続され、

15 前記標準モデル作成装置はさらに、

前記標準モデル作成手段が作成した標準モデルを前記端末装置に送信する標準モデル送信手段を備える

ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

20 1 3. 前記参照モデル及び前記標準モデルは、1 以上のガウス分布を用いて表現され、

前記参照モデル記憶手段は、少なくともガウス分布の混合数が異なる 1 対の参照モデルを記憶し、

前記統計量推定部は、前記 1 対の参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算する

25

ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

1 4 . 前記標準モデル作成手段はさらに、

外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、新たな参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段を備える

ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

1 5 . 前記参照モデル準備手段は、さらに、前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行う

ことを特徴とする請求の範囲 1 4 記載の標準モデル作成装置。

1 6 . 前記参照モデル準備手段は、認識の対象に関する情報である利用情報、及び作成する標準モデルの仕様に関する情報である仕様情報の少なくとも一方に基づいて、前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行う

ことを特徴とする請求の範囲 1 5 記載の標準モデル作成装置。

1 7 . 前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報、及び、音声認識の対象となる属性に関する情報である利用情報の少なくとも一方と、前記参照モデル記憶手段に記憶された参照モデルとに基づいて、前記利用情報及び前記仕様情報の少なくとも一方と前記参照モデルとの類似度を示す類似度情報を作成する類似度情報作成手段を備え、

25 前記参照モデル準備手段は、前記類似度情報作成手段が作成した類似度情報に基づいて、前記参照モデル記憶手段が記憶する参照モデルの更

新及び追加の少なくとも一方を行うか否かを決定する

ことを特徴とする請求の範囲 15 記載の標準モデル作成装置。

18. 前記初期標準モデル作成部は、前記統計量推定部が標準モデルの統計量を計算するために用いる、1 以上の前記参照モデルを用いて前記標準モデルを特定する統計量の初期値を決定する

ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

19. 前記初期標準モデル作成部は、標準モデルの種類を識別するクラス ID に基づいて、前記初期値を決定する

ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

20. 前記初期標準モデル作成部は、前記参照モデルから前記クラス ID を特定し、特定したクラス ID に対応づけられた初期値を前記初期値と決定する

ことを特徴とする請求の範囲 19 記載の標準モデル作成装置。

21. 前記初期標準モデル作成部は、前記クラス ID と前記初期値と前記参照モデルとの対応を示す対応表を保持し、前記対応表に従って、前記初期値を決定する

ことを特徴とする請求の範囲 20 記載の標準モデル作成装置。

22. 前記初期標準モデル作成部は、前記クラス ID が対応づけられた初期値であるクラス ID 付き初期標準モデル、又は、前記クラス ID が対応づけられた参照モデルであるクラス ID 付き参照モデルを作成又は外部から取得することによって、前記対応表を生成する

ことを特徴とする請求の範囲 2 1 記載の標準モデル作成装置。

2 3 . 前記参照モデル記憶手段は、複数の参照モデルを記憶し、

前記統計量推定部は、前記参照モデル記憶手段に記憶された複数の参
5 照モデルに対して重み付けられた前記確率又は尤度を最大化又は極大化
するように前記統計量を計算する

ことを特徴とする請求の範囲 1 記載の標準モデル作成装置。

2 4 . 音声の特徴を示す周波数のパラメータを出力確率で表現する確率
10 モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標
準モデルを作成する方法であって、

一定の属性を有する音声の特徴を示す確率モデルである 1 以上の参照
モデルを記憶する参照モデル記憶手段から 1 以上の参照モデルを読み出
す参照モデル読み出しステップと、

15 読み出された参照モデルの統計量を用いて前記標準モデルの統計量を
計算することによって標準モデルを作成する標準モデル作成ステップと
を含み、

前記標準モデル作成ステップは、

作成する標準モデルの構造を決定する標準モデル構造決定サブステッ
20 プと、

構造が決定された標準モデルを特定する統計量の初期値を決定する初
期標準モデル作成サブステップと、

初期値が決定された標準モデルの前記参照モデルに対する確率又は尤
度を最大化又は極大化するように前記標準モデルの統計量を推定して計
25 算する統計量推定サブステップとを有する

ことを特徴とする標準モデル作成方法。

25. 音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置のためのプログラムであって、

- 5 一定の属性を有する音声の特徴を示す確率モデルである1以上の参照モデルを記憶する参照モデル記憶手段から1以上の参照モデルを読み出す参照モデル読み出しステップと、

読み出された参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成ステップと

- 10 を含み、

前記標準モデル作成ステップは、

作成する標準モデルの構造を決定する標準モデル構造決定サブステップと、

- 15 構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成サブステップと、

初期値が決定された標準モデルの前記参照モデルに対する確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定サブステップとを有する

ことを特徴とするプログラム。

図1

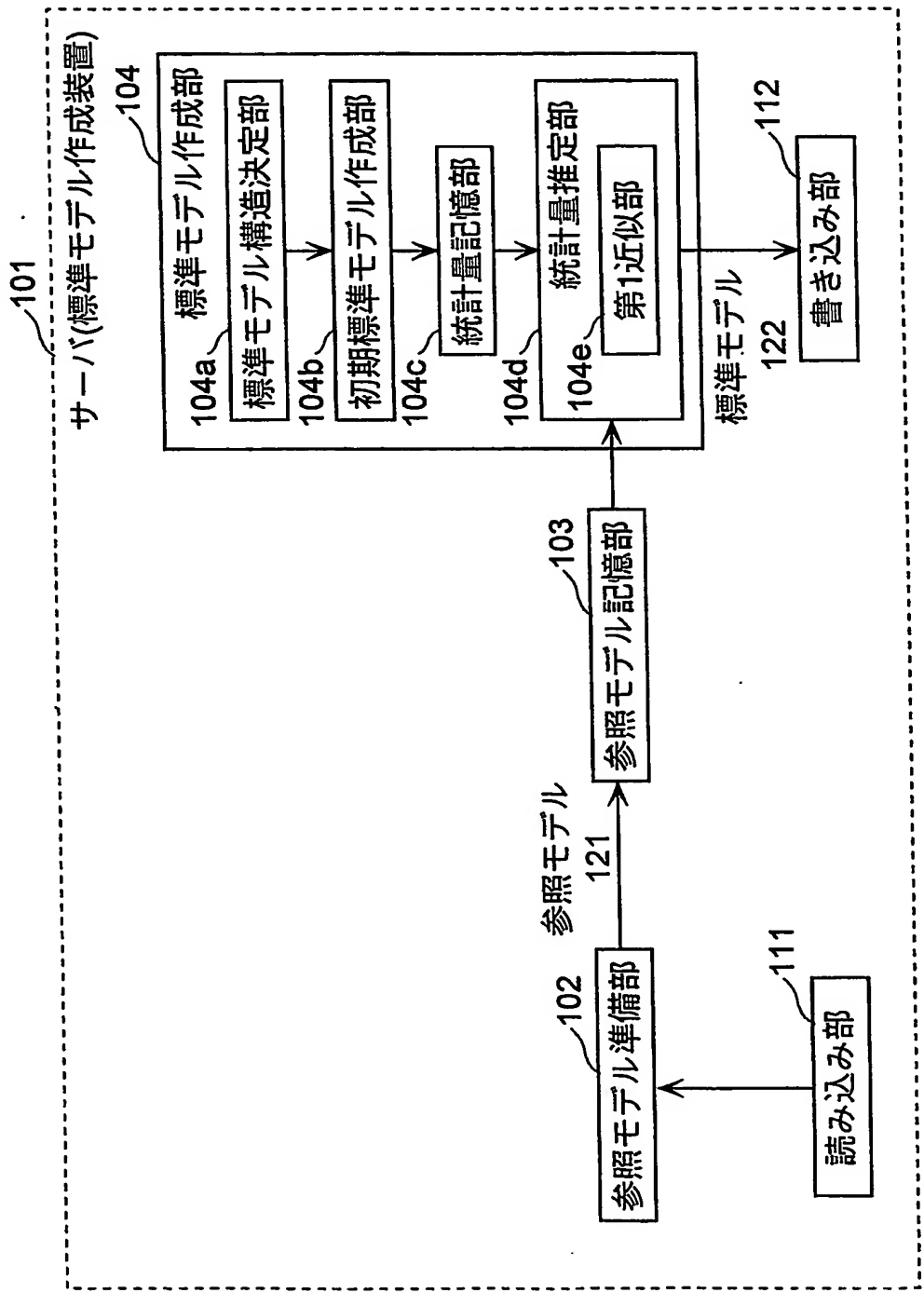


図2

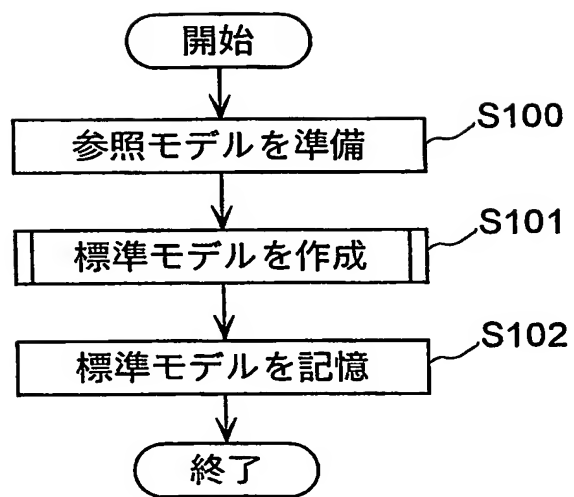
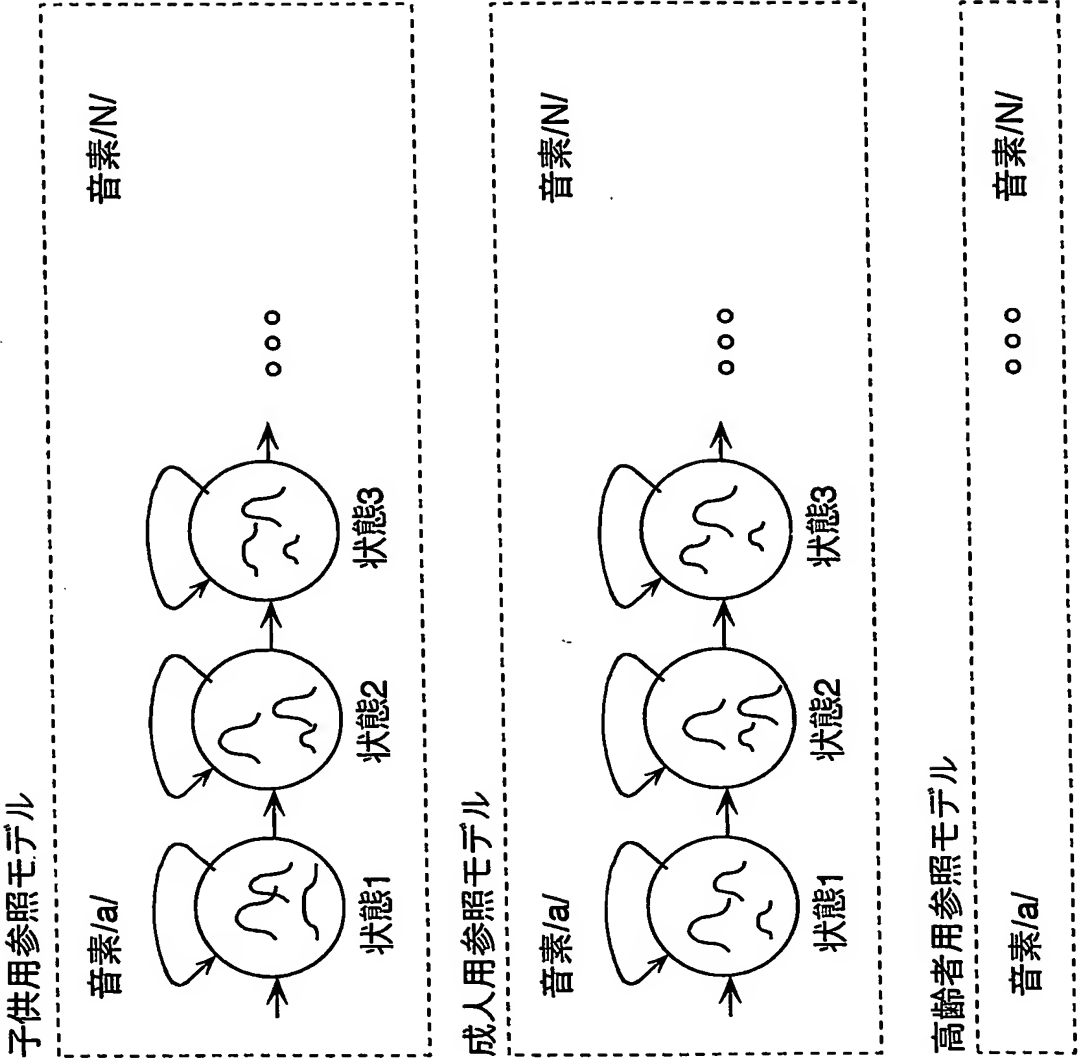


図3



参照モデル
121

図4

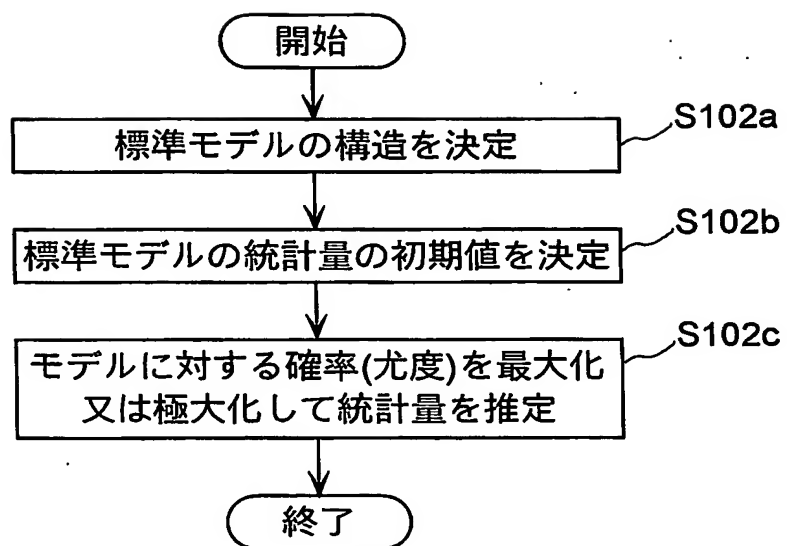


図5

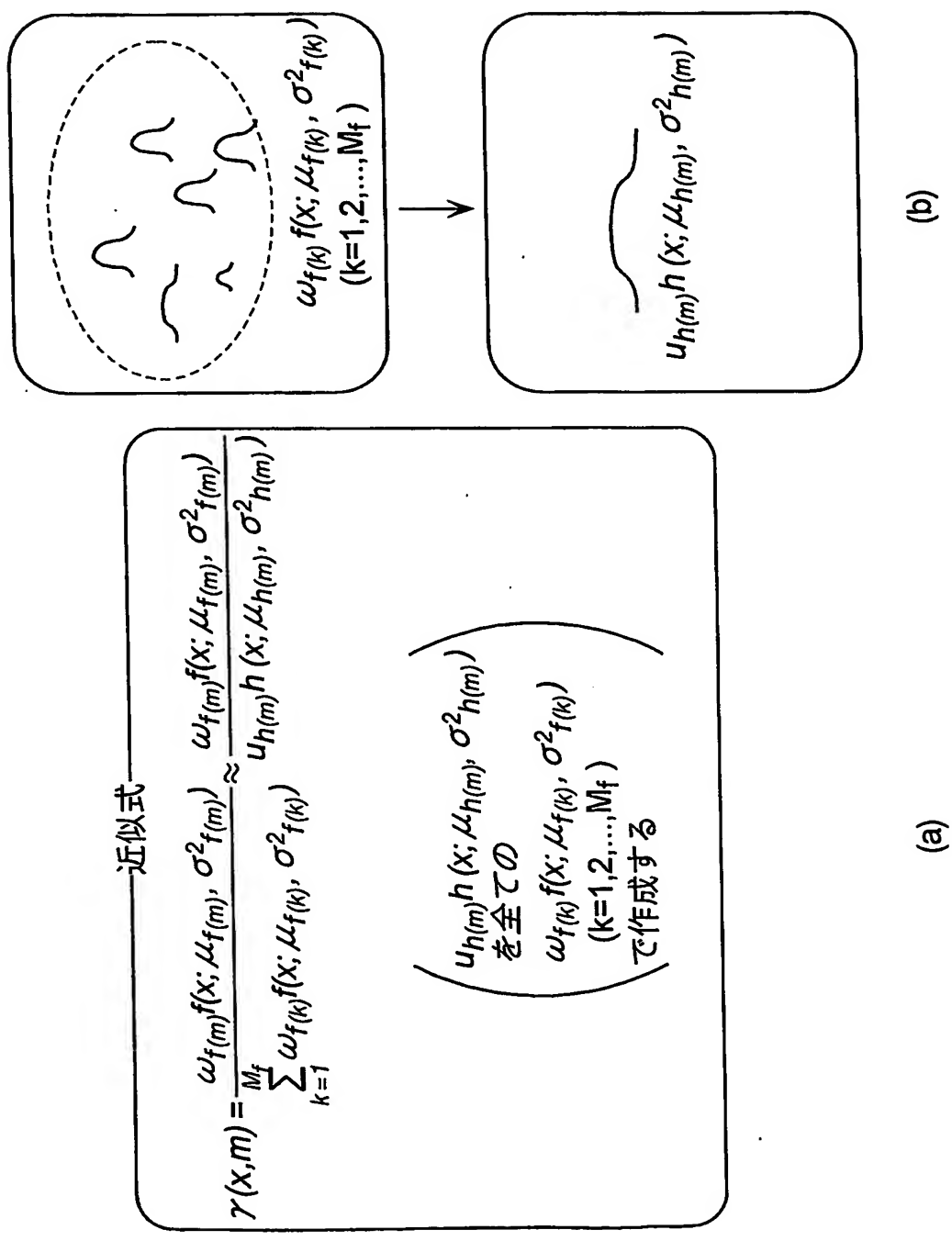


図6

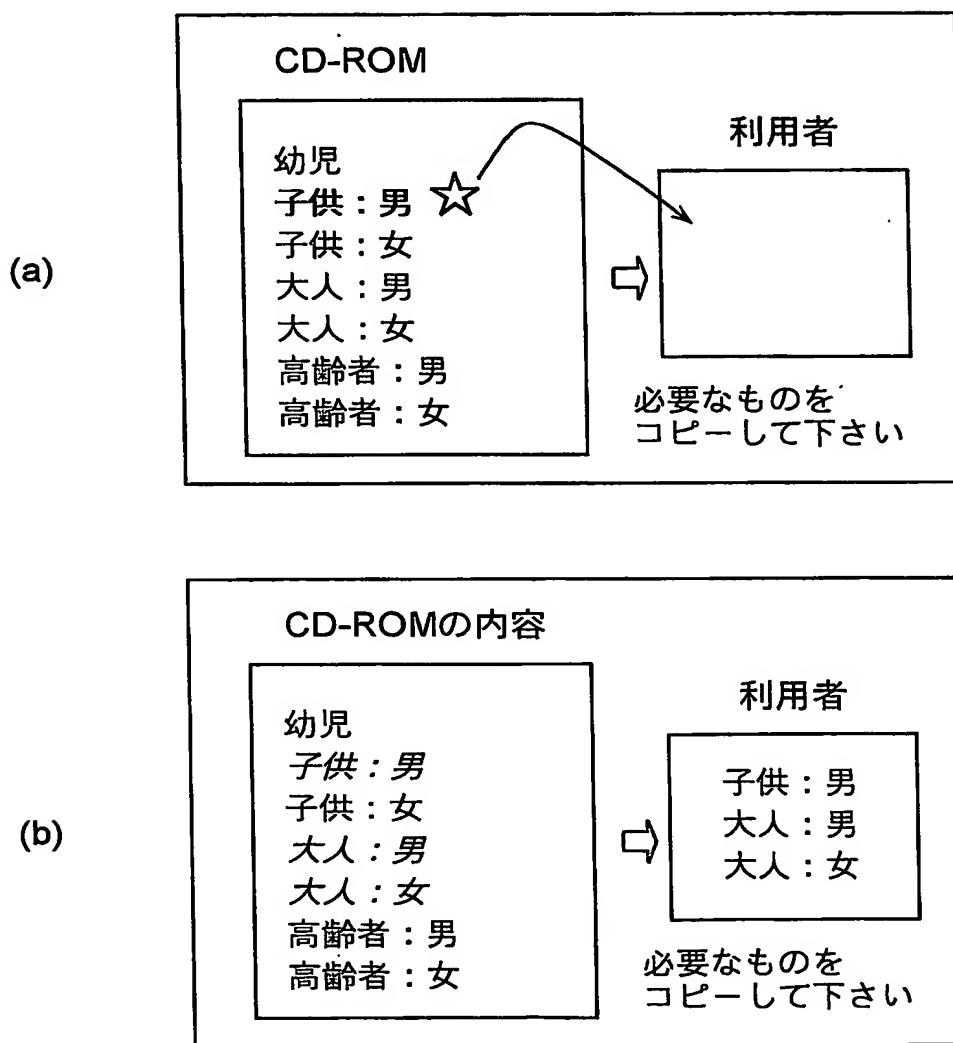


図7

(a)

モデルの複雑さ(混合分布数)

3個

10個

20個

(b)

利用機器

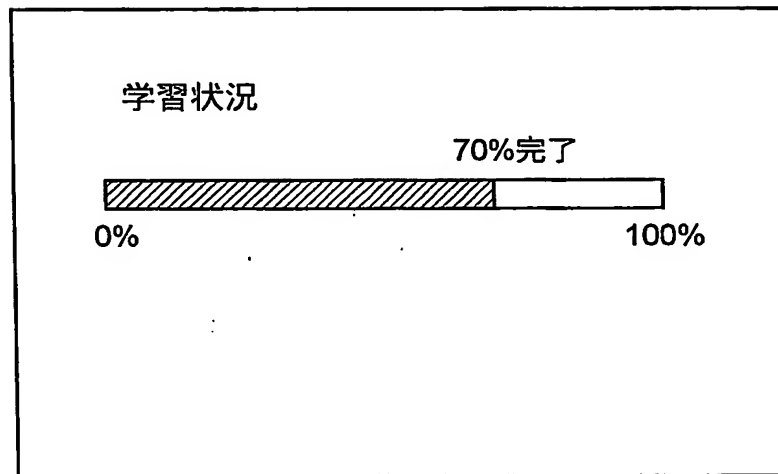
テレビ用

カーナビ用

携帯電話用

図8

(a)



(b)

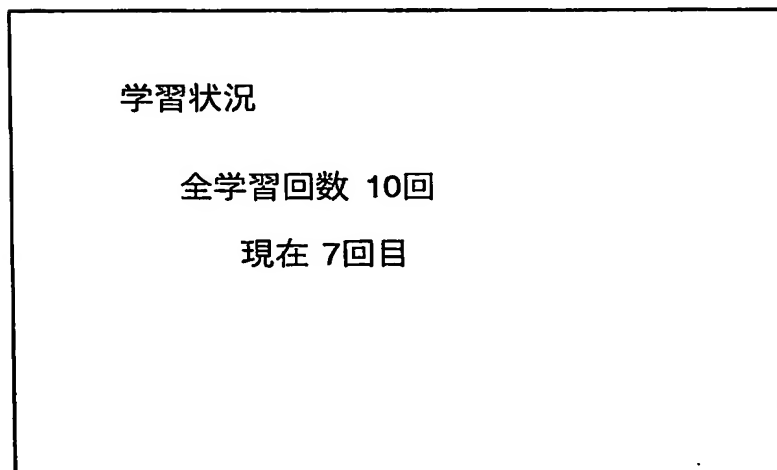


図9

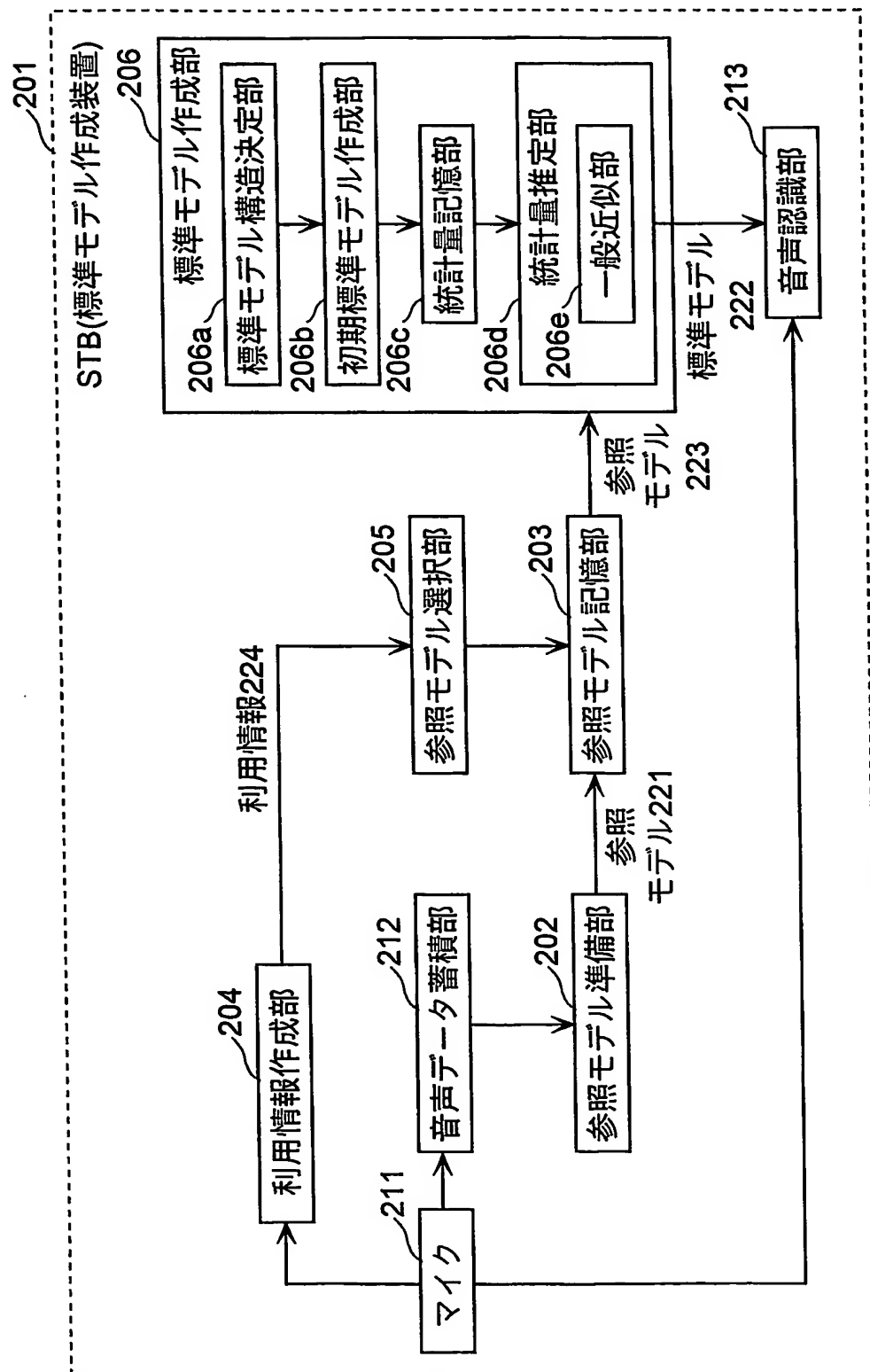


図10

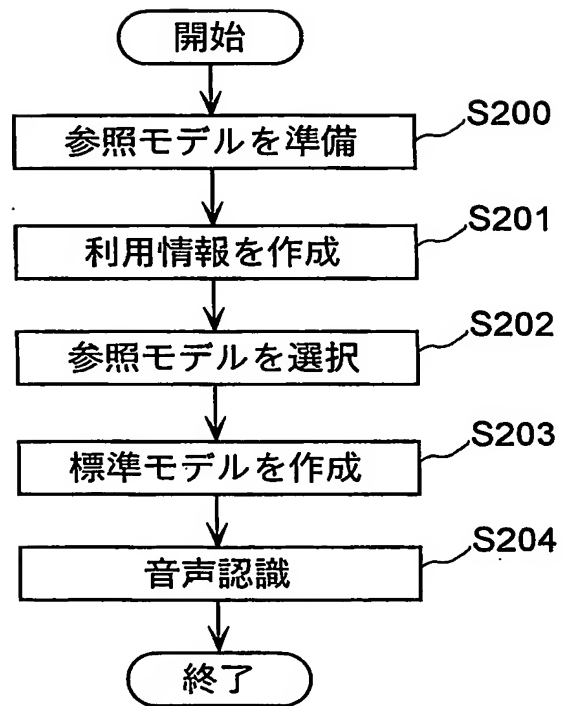
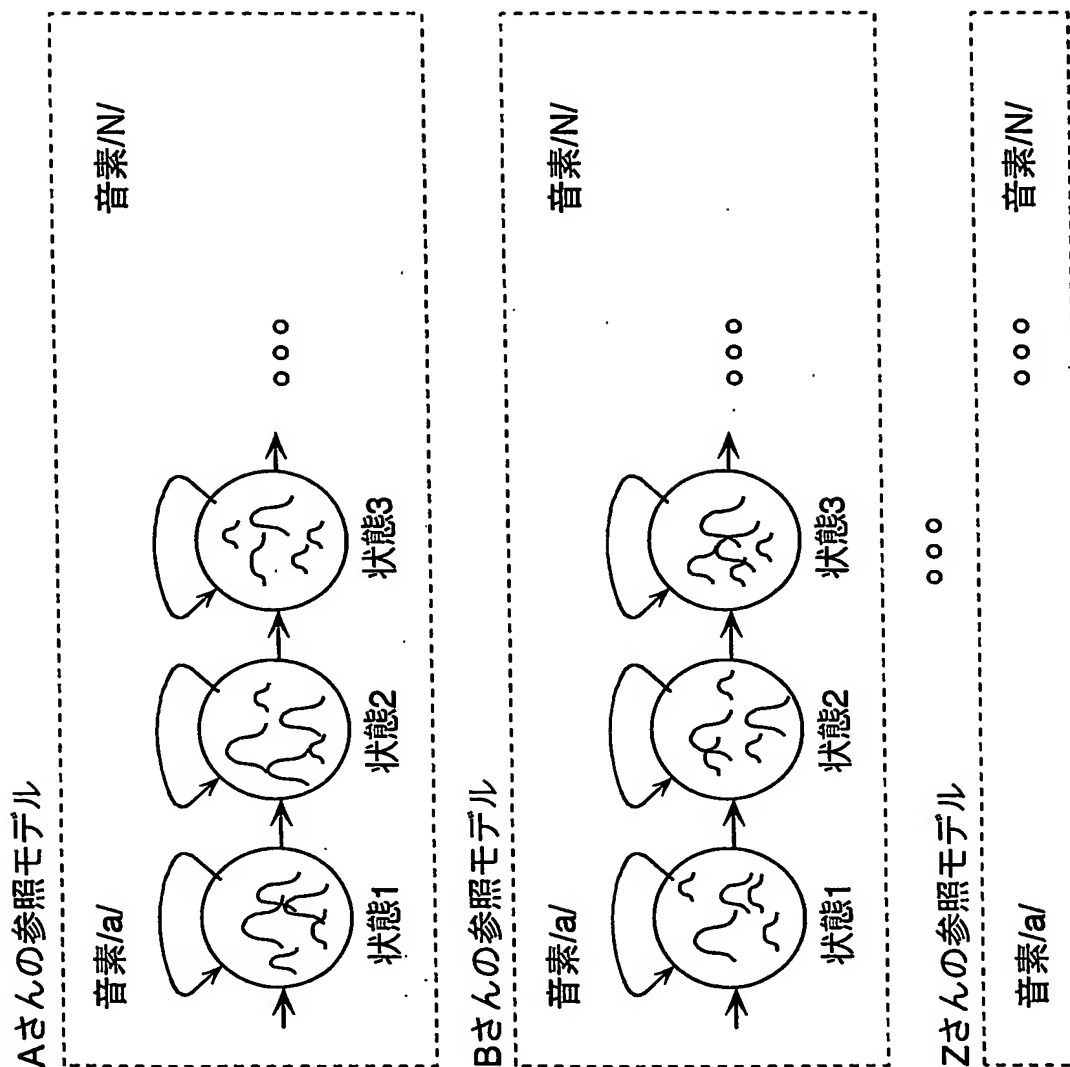


図11



参照モデル
221

図12

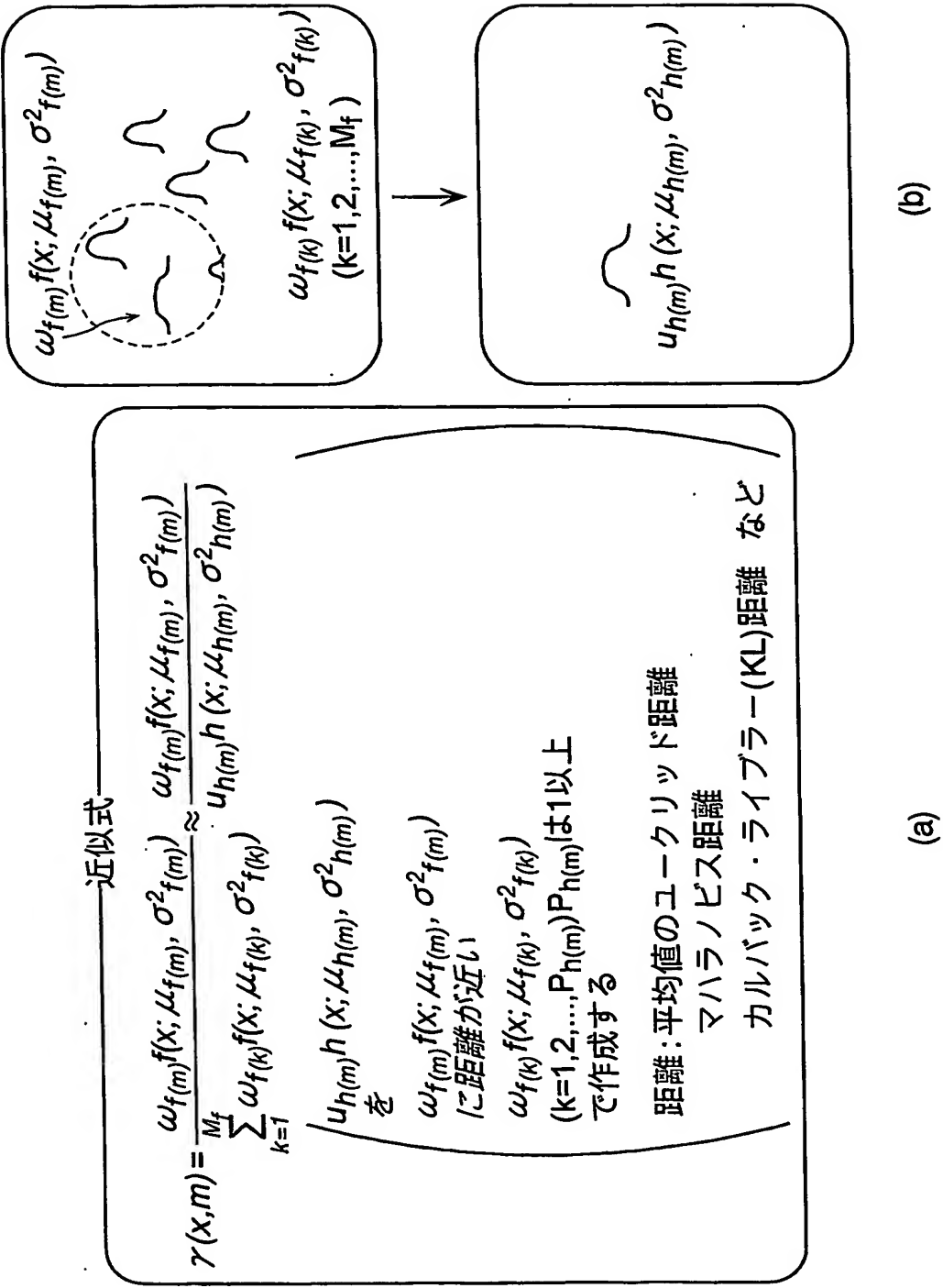


図14

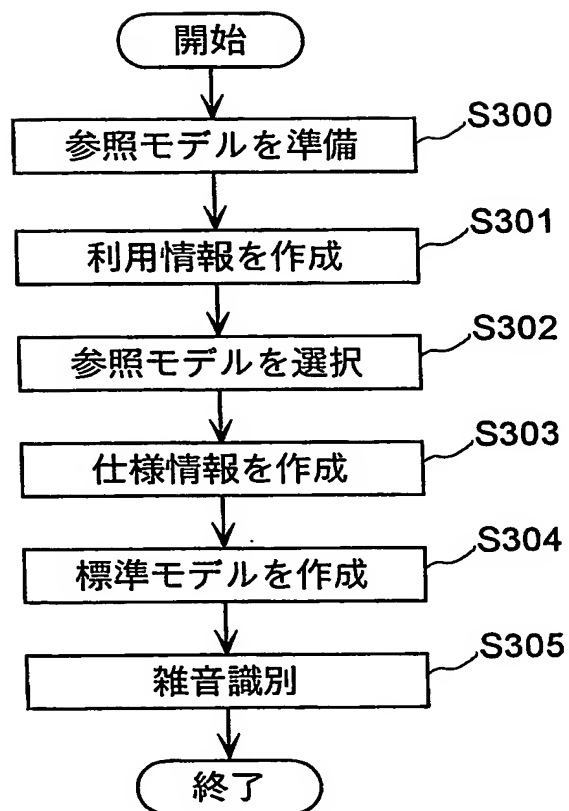


図15

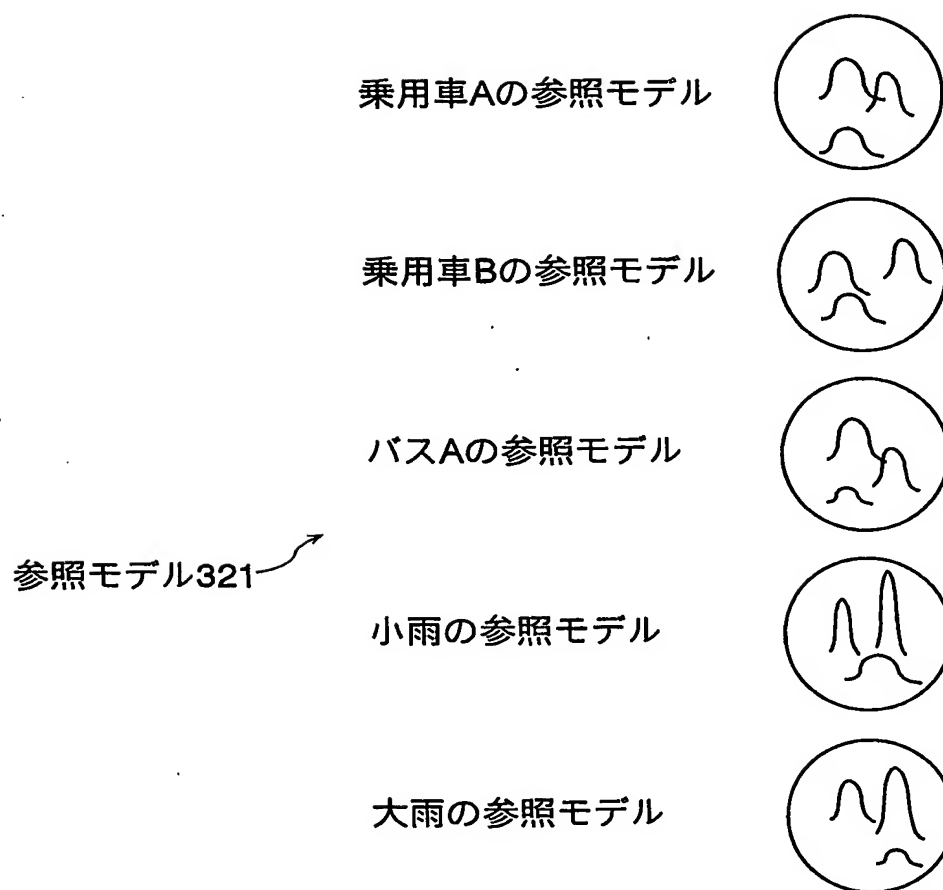
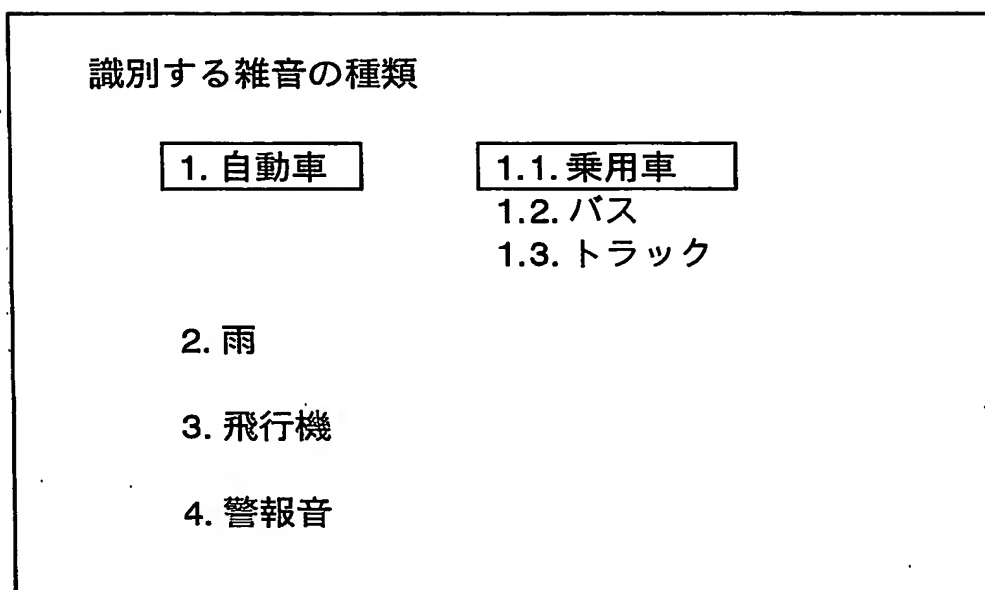


図16



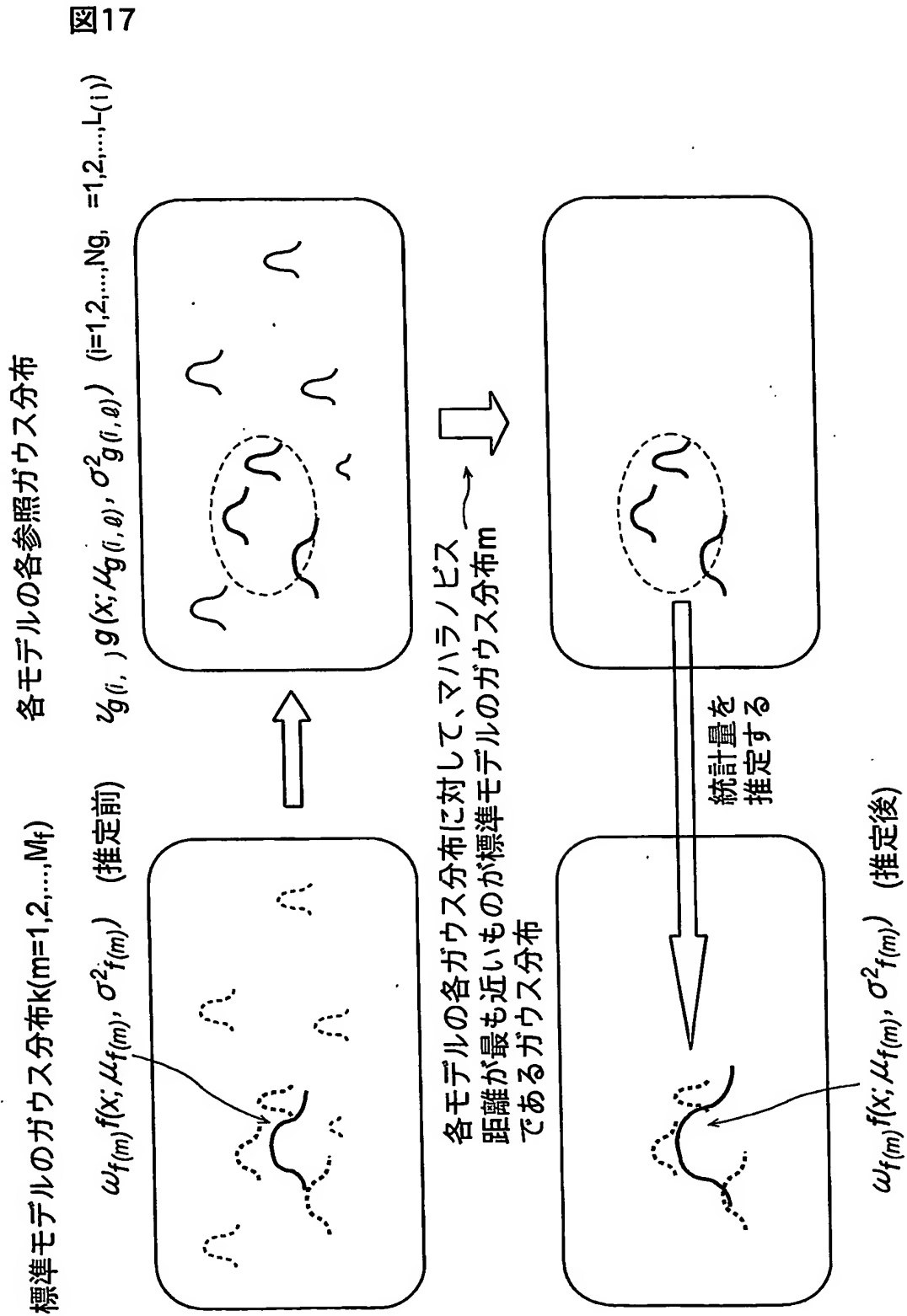


図 17

図18

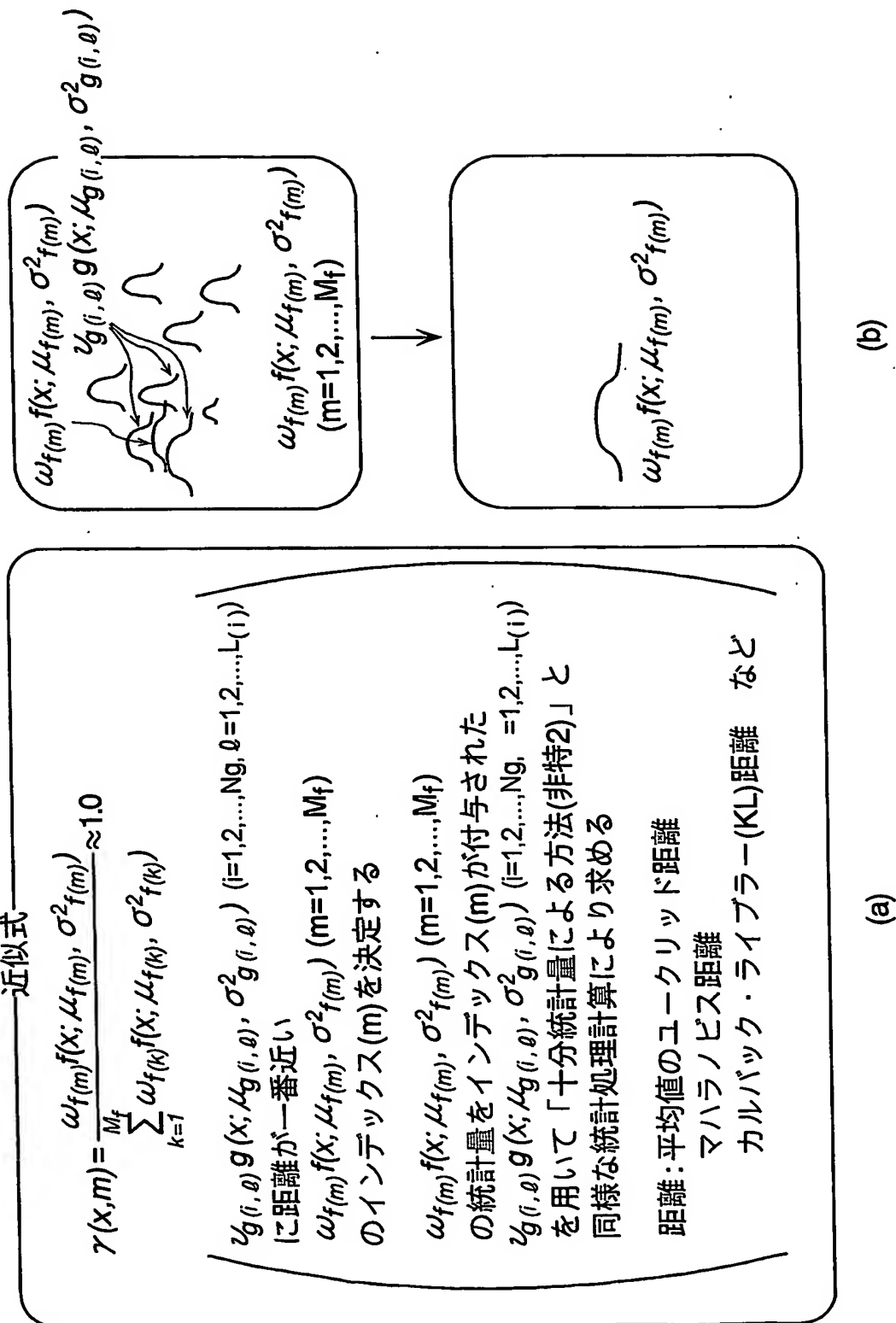


図19

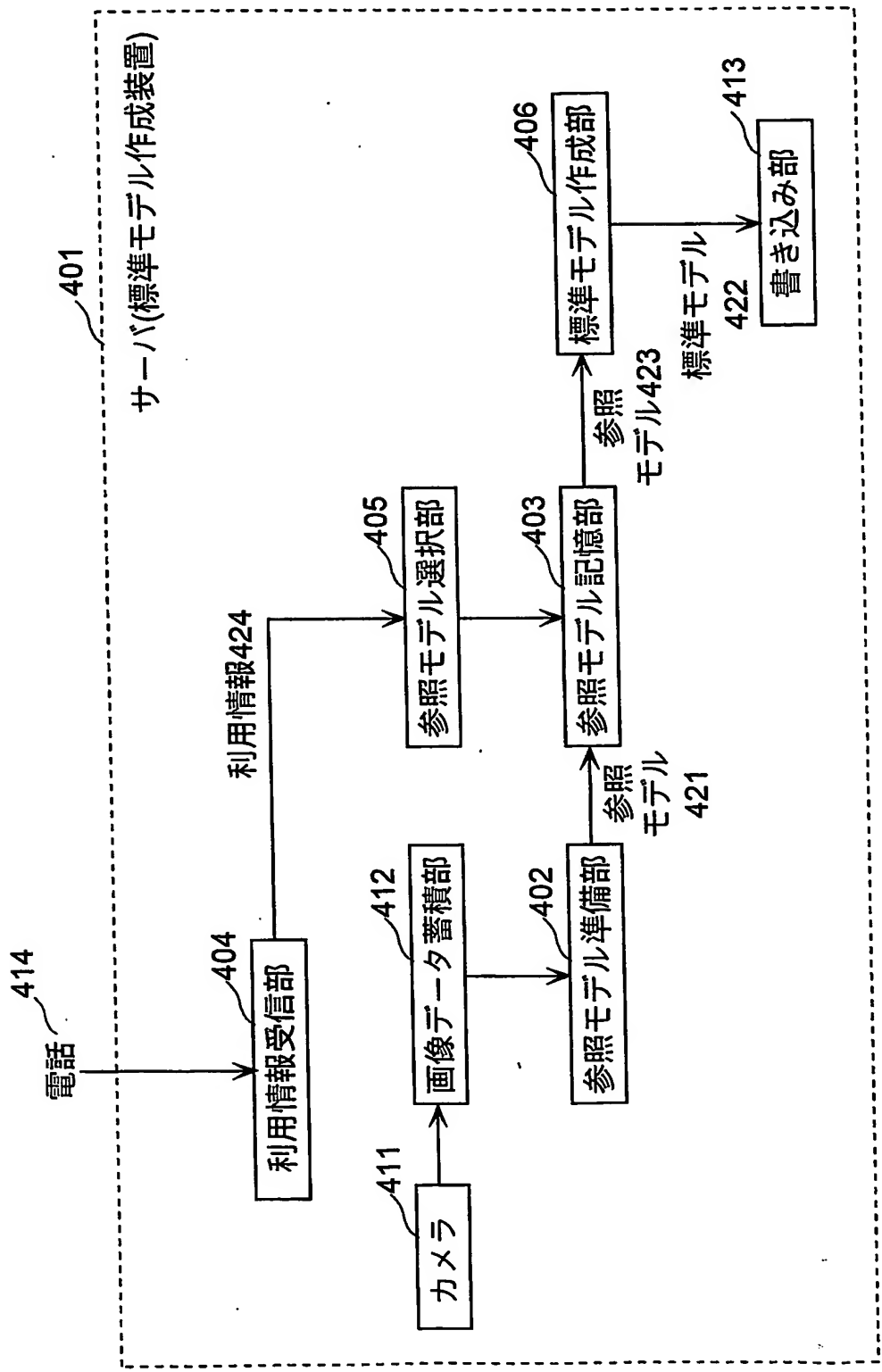


図20

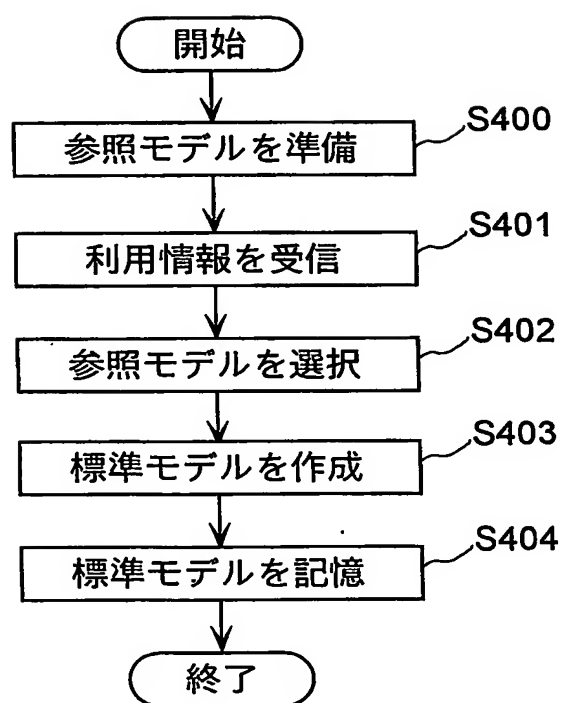


図21

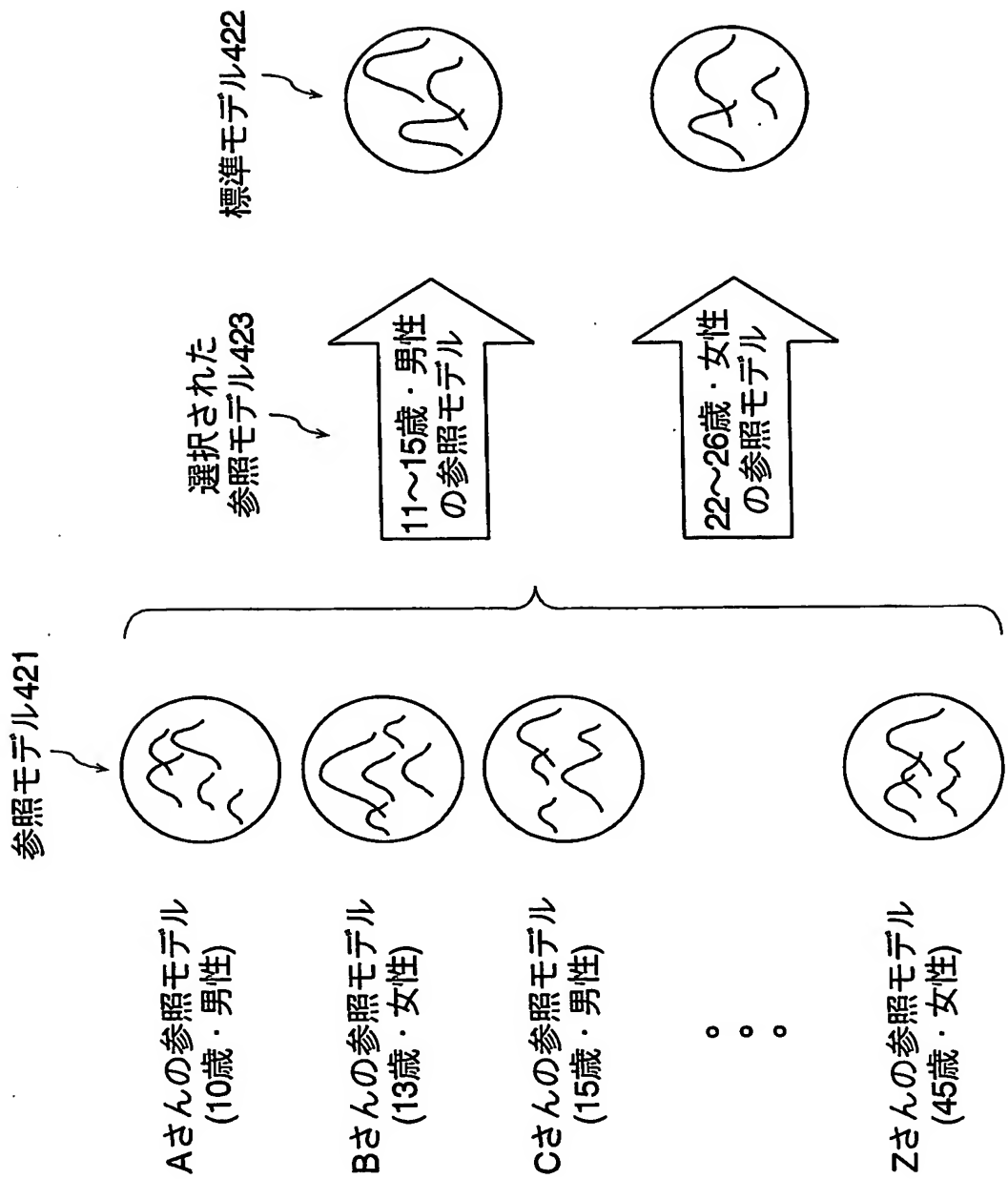


図22

名前「お父さん」	生別 <input checked="" type="checkbox"/> 男 <input type="checkbox"/> 女
住所「大阪市」	年齢 「50歳」
趣味 <input checked="" type="checkbox"/> ドライブ	<input type="checkbox"/> マッサージ
<input type="checkbox"/> スポーツ観戦	<input type="checkbox"/> コンピュータ
<input type="checkbox"/> 釣り	<input type="checkbox"/> ゲーム
<input type="checkbox"/> ショッピング	
<input checked="" type="checkbox"/> 温泉	
<input type="checkbox"/> ゴルフ	
<input type="checkbox"/> キャンプ	

図23

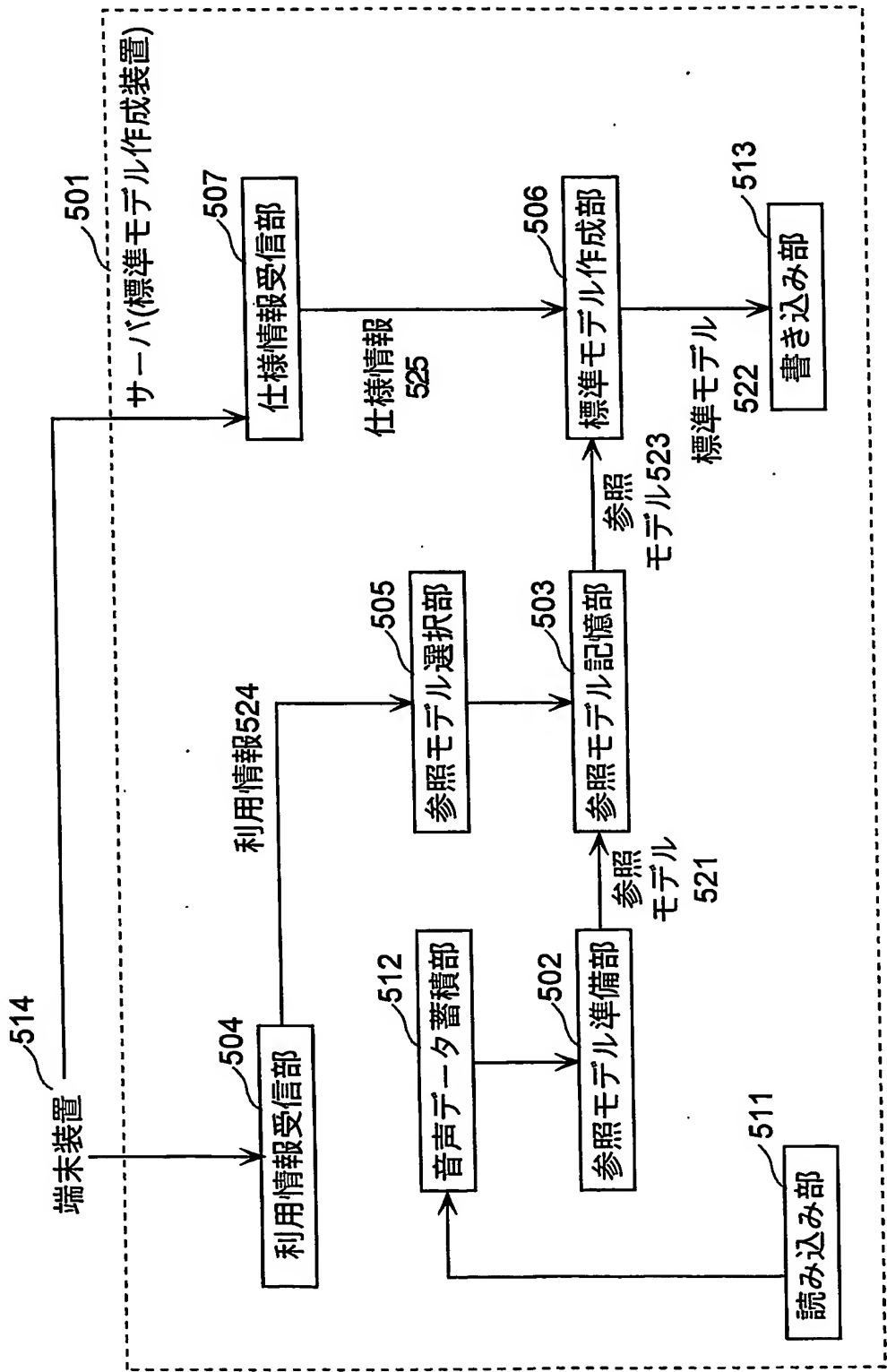


図24

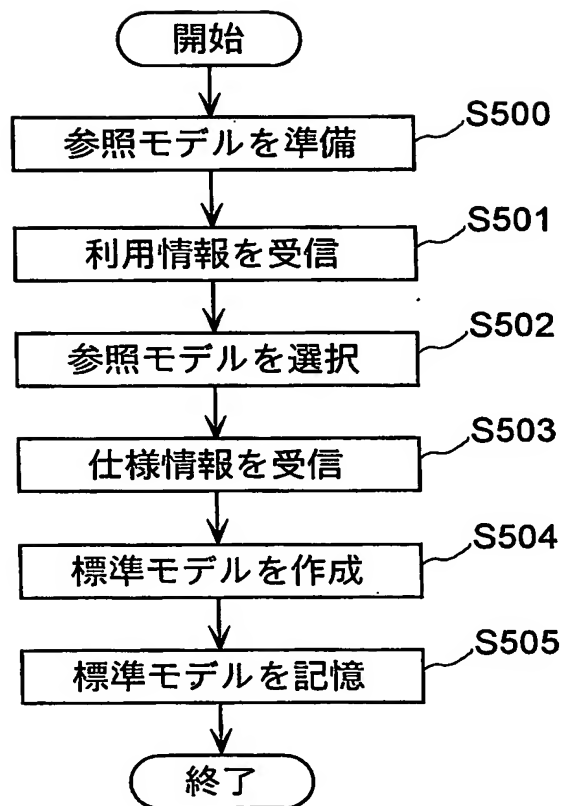


図25

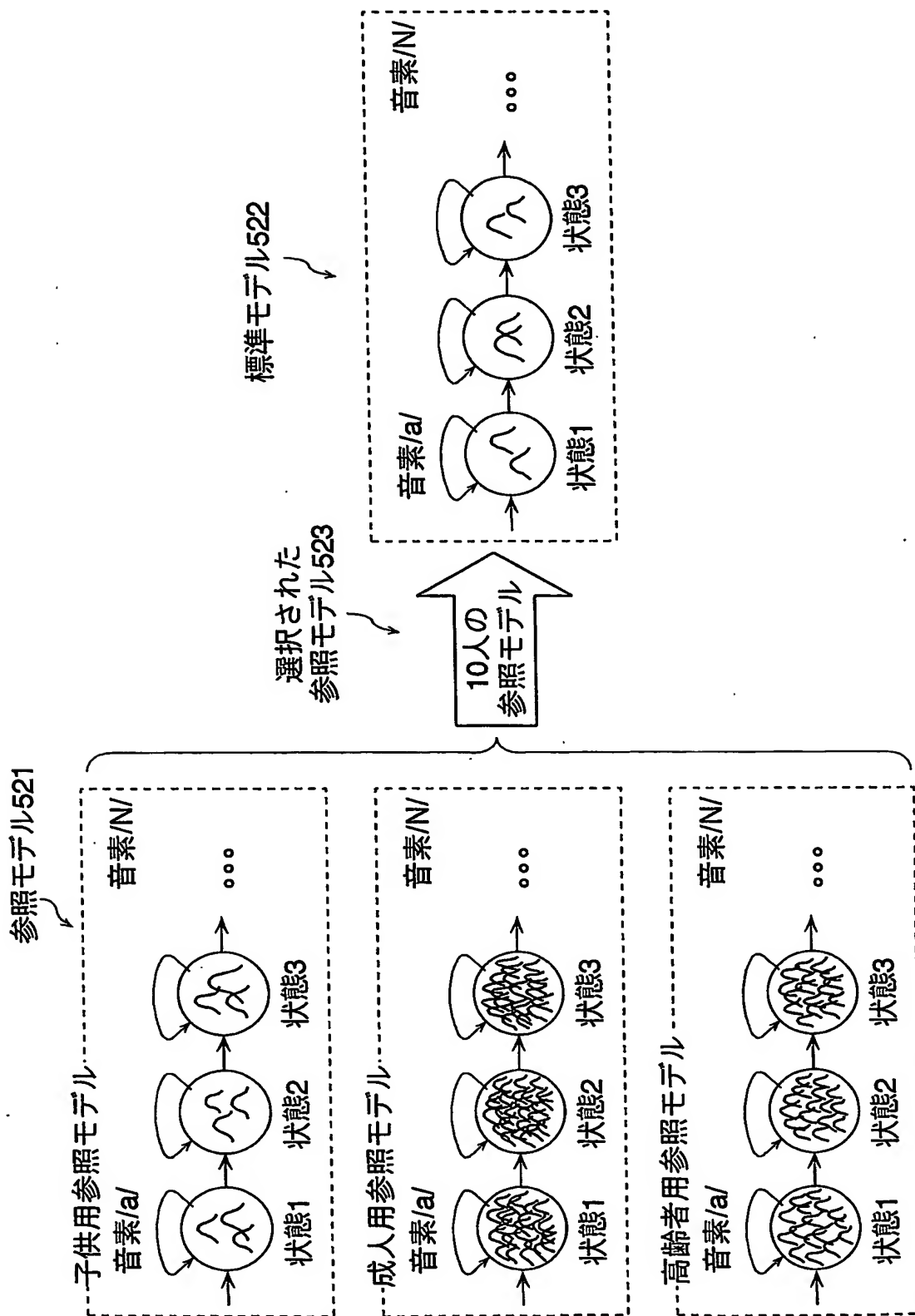


図26

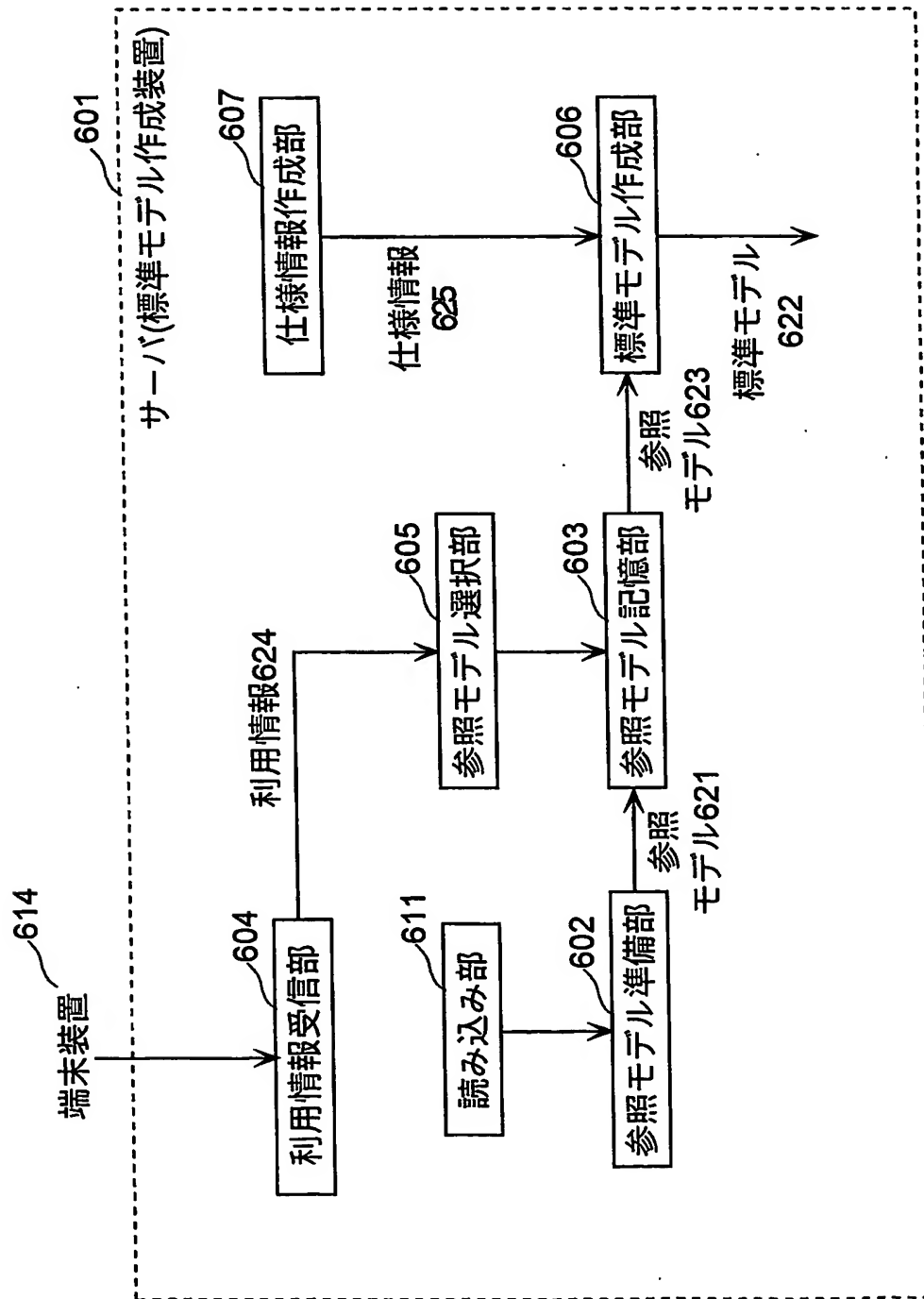


図27

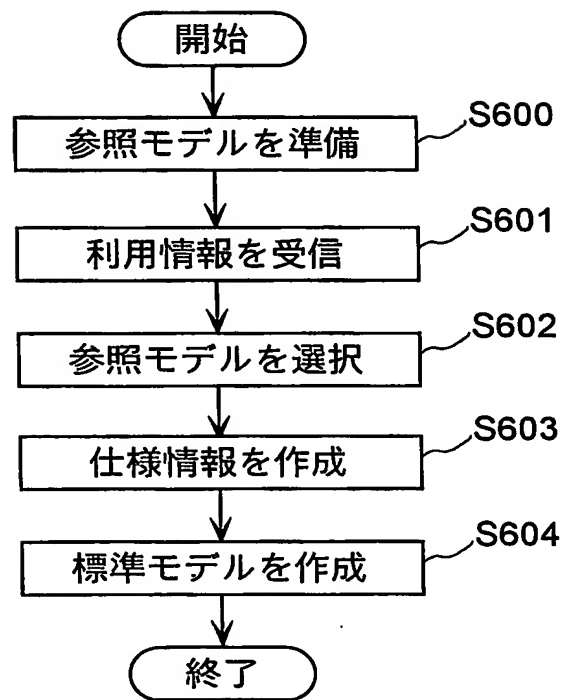


図28

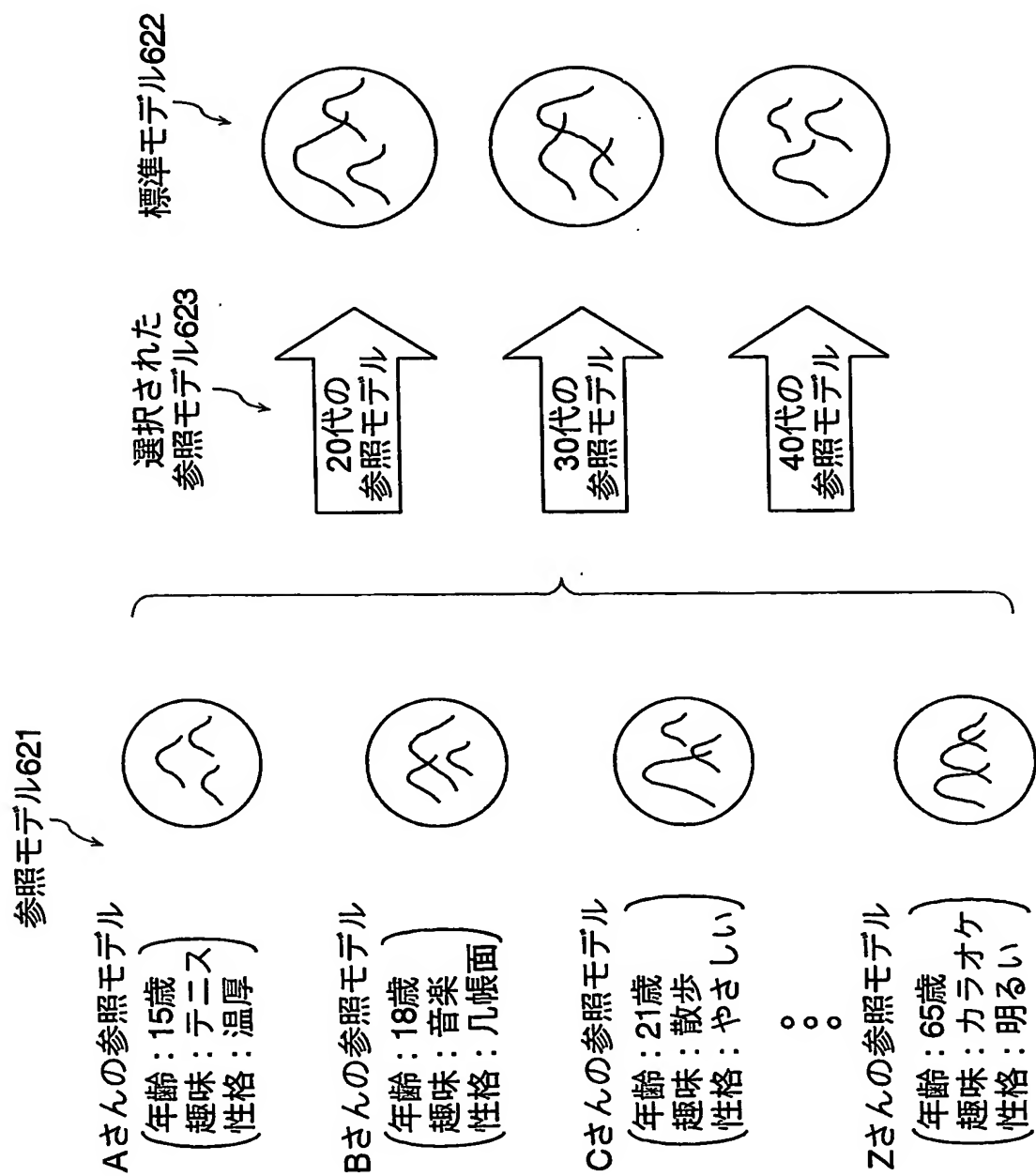


図29

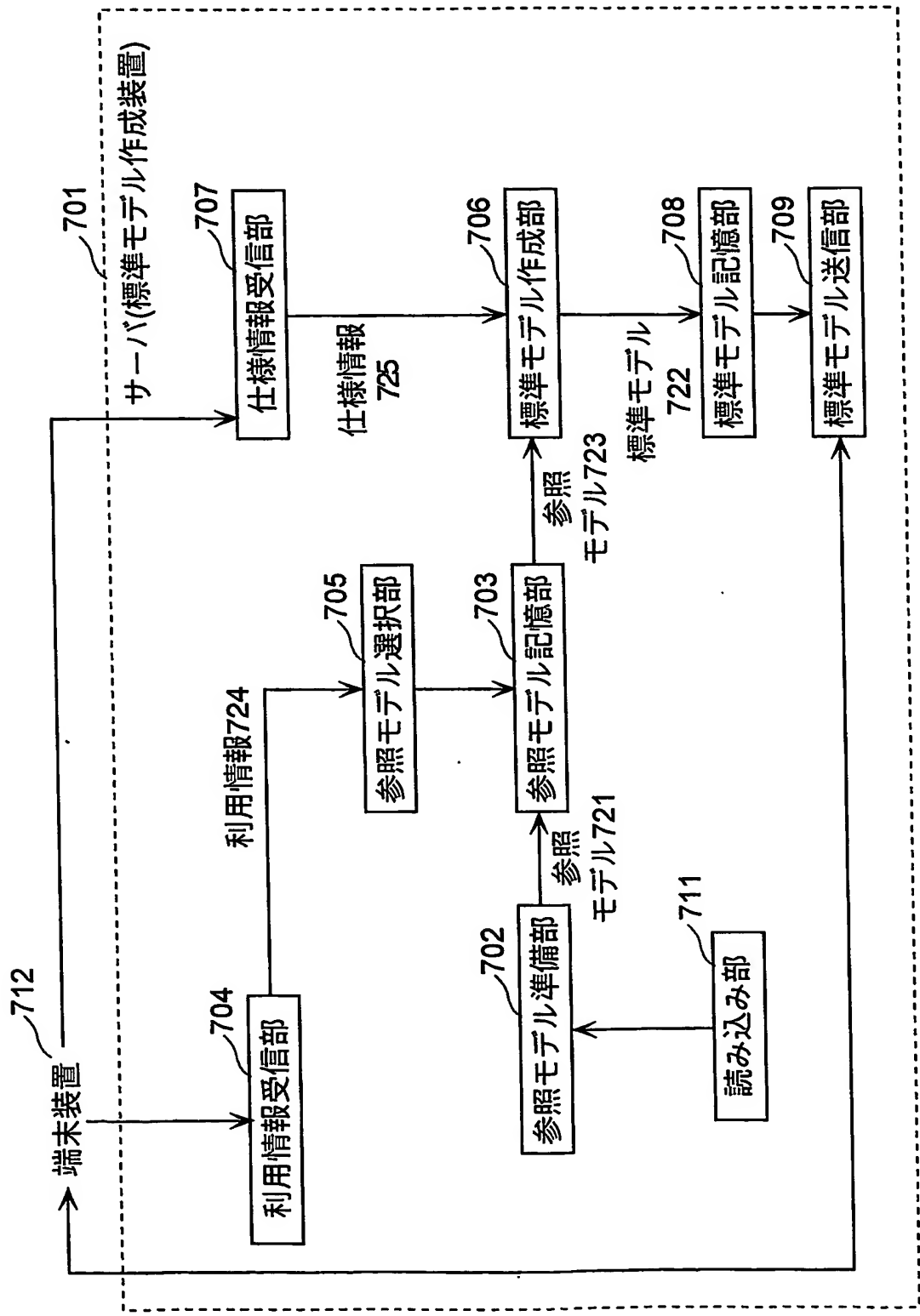


図30

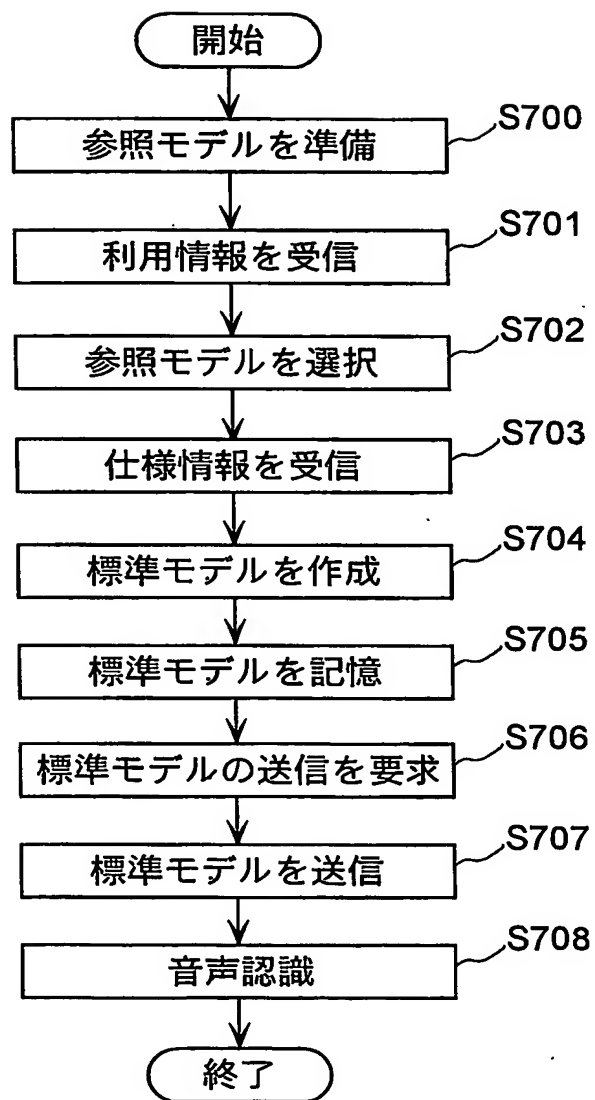


図31

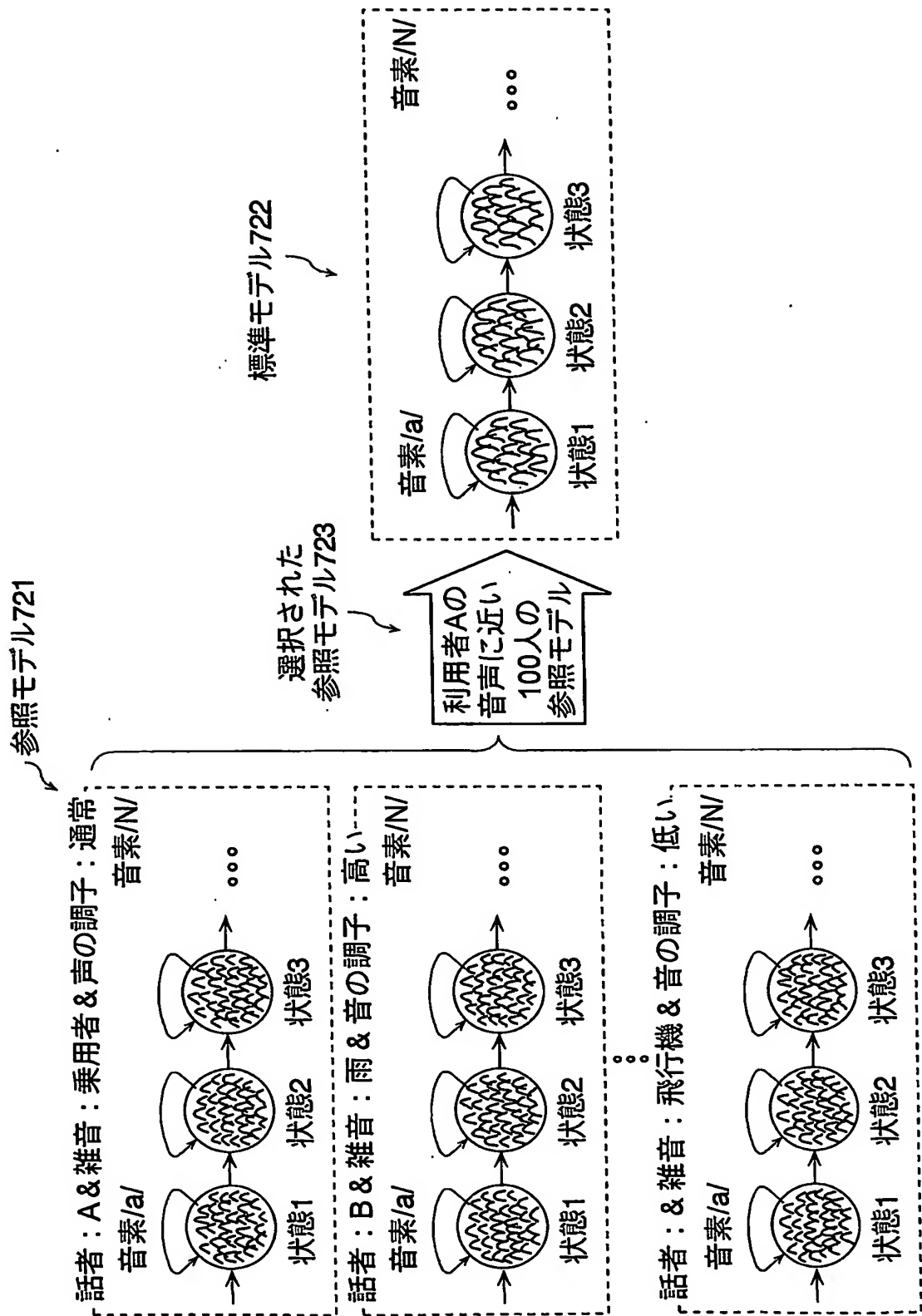


図32

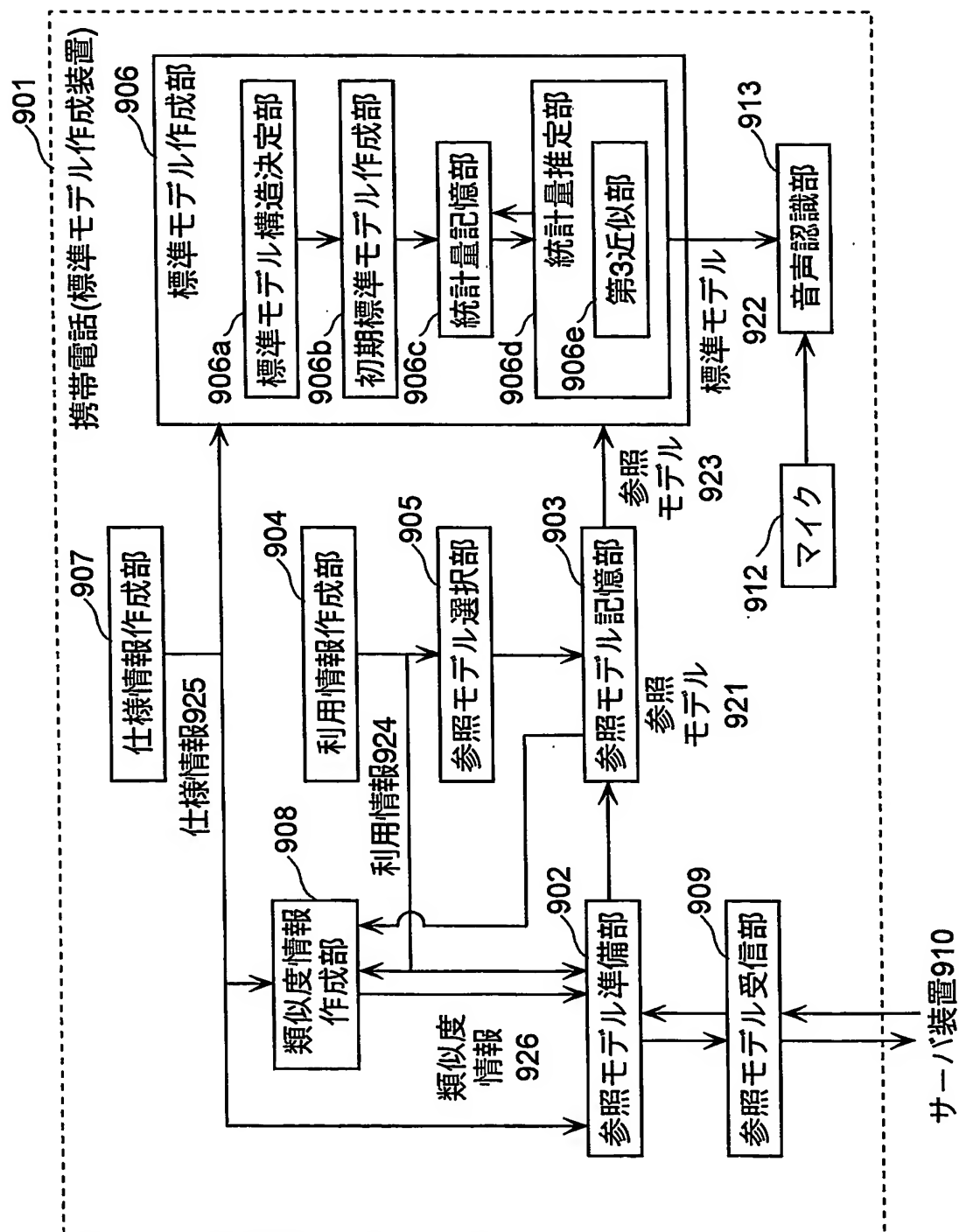


図33

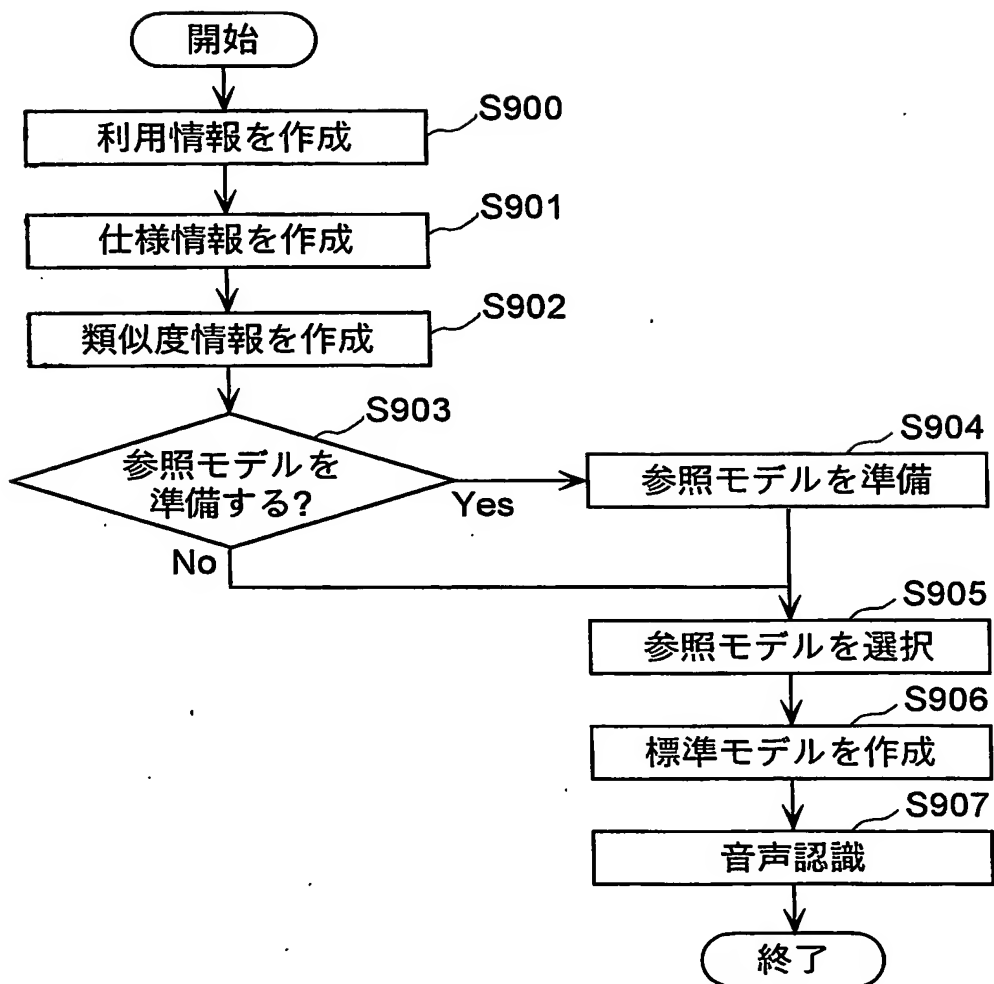


図34

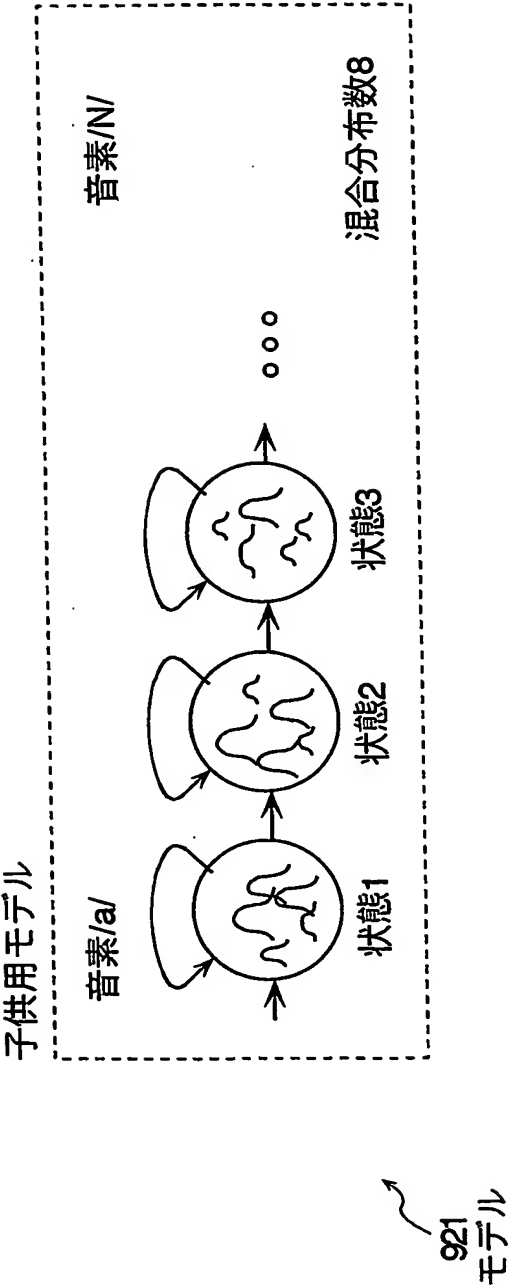
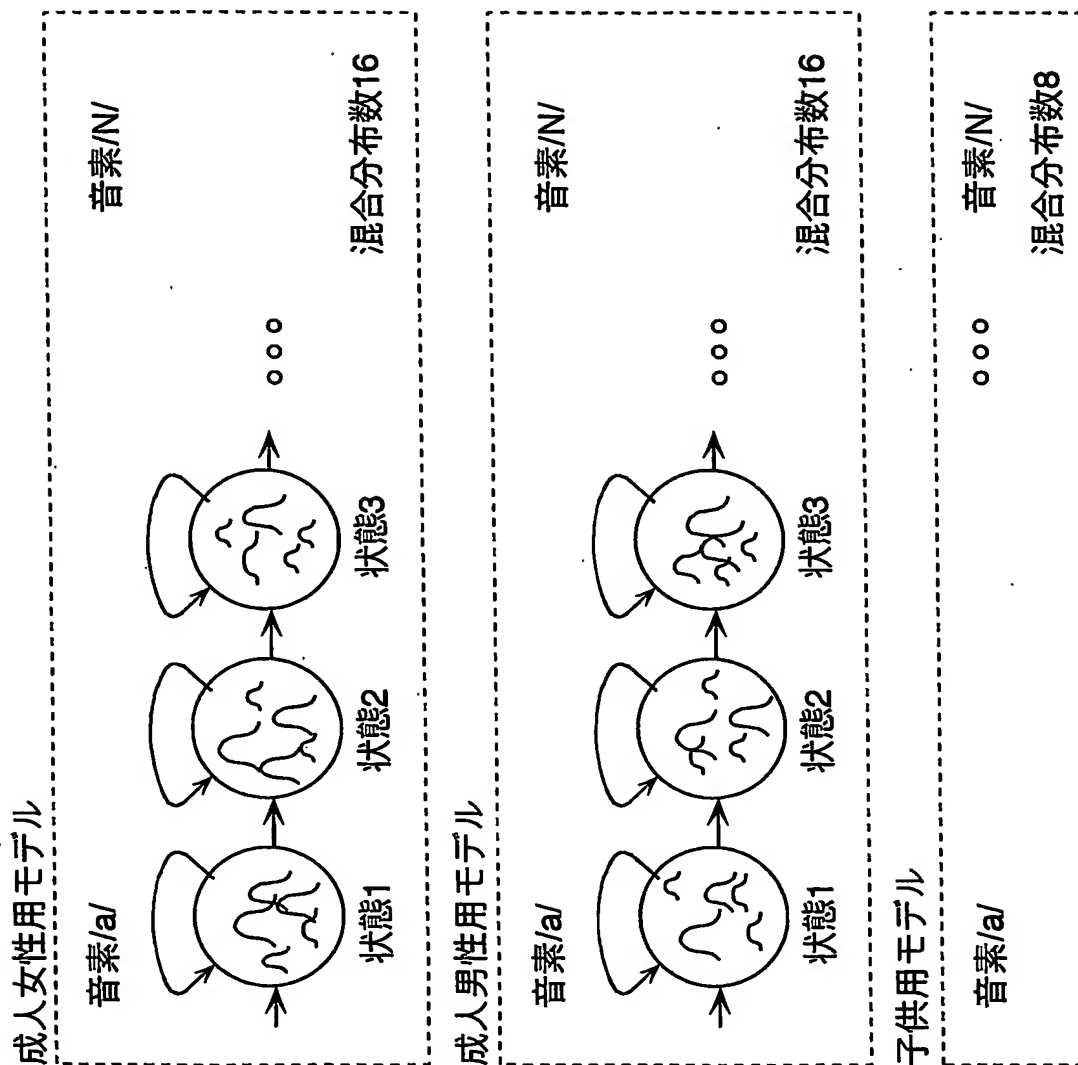


図35



921
モデル

図36

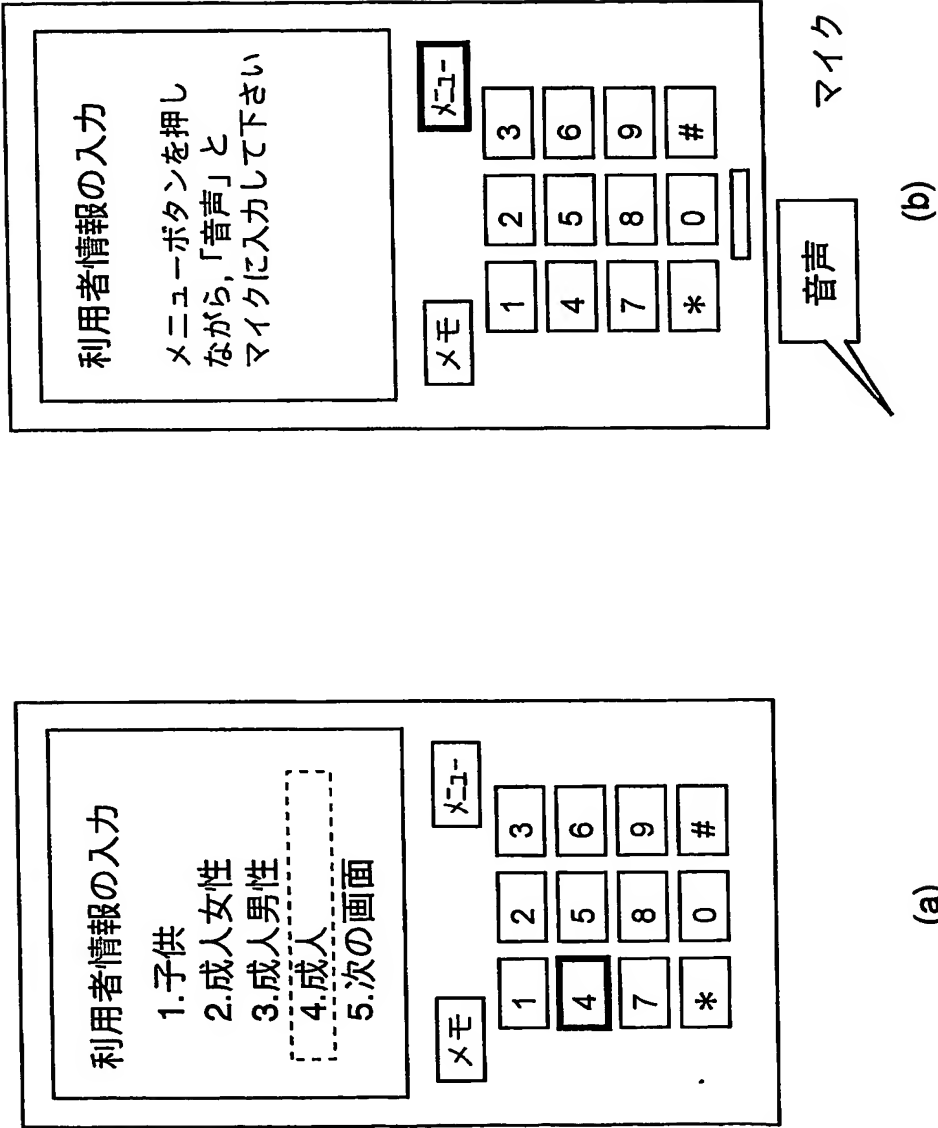
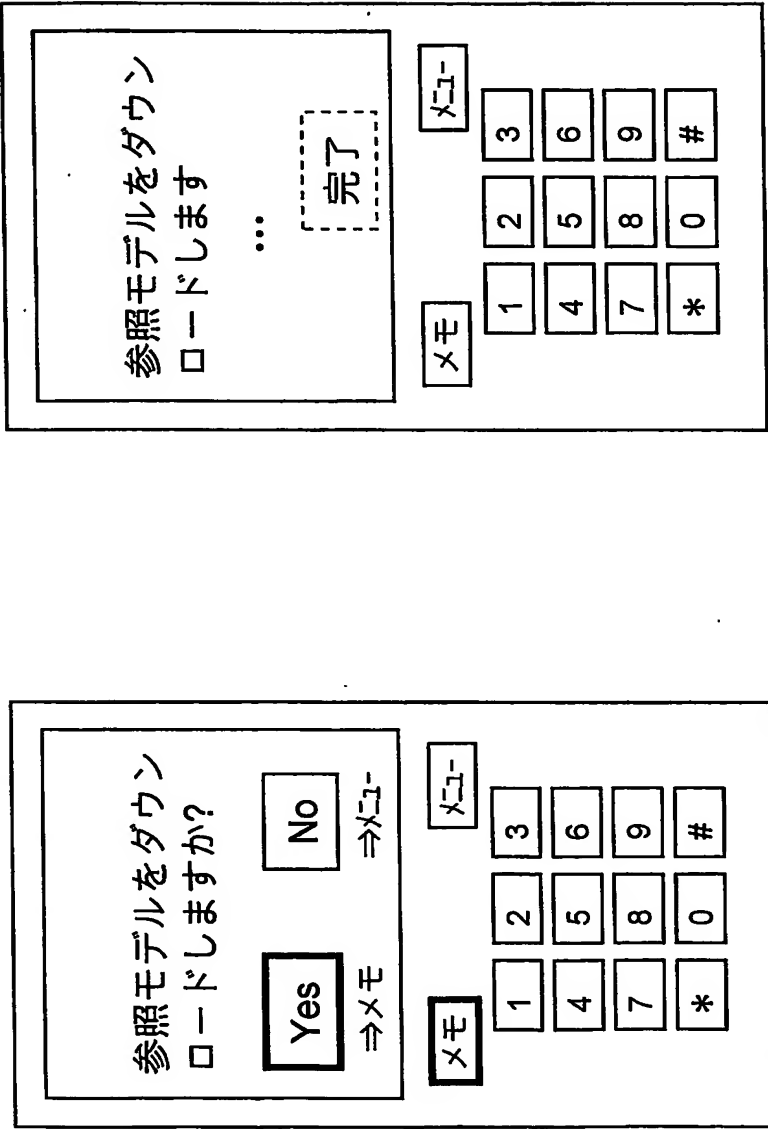


図37



(b)

(a)

図38

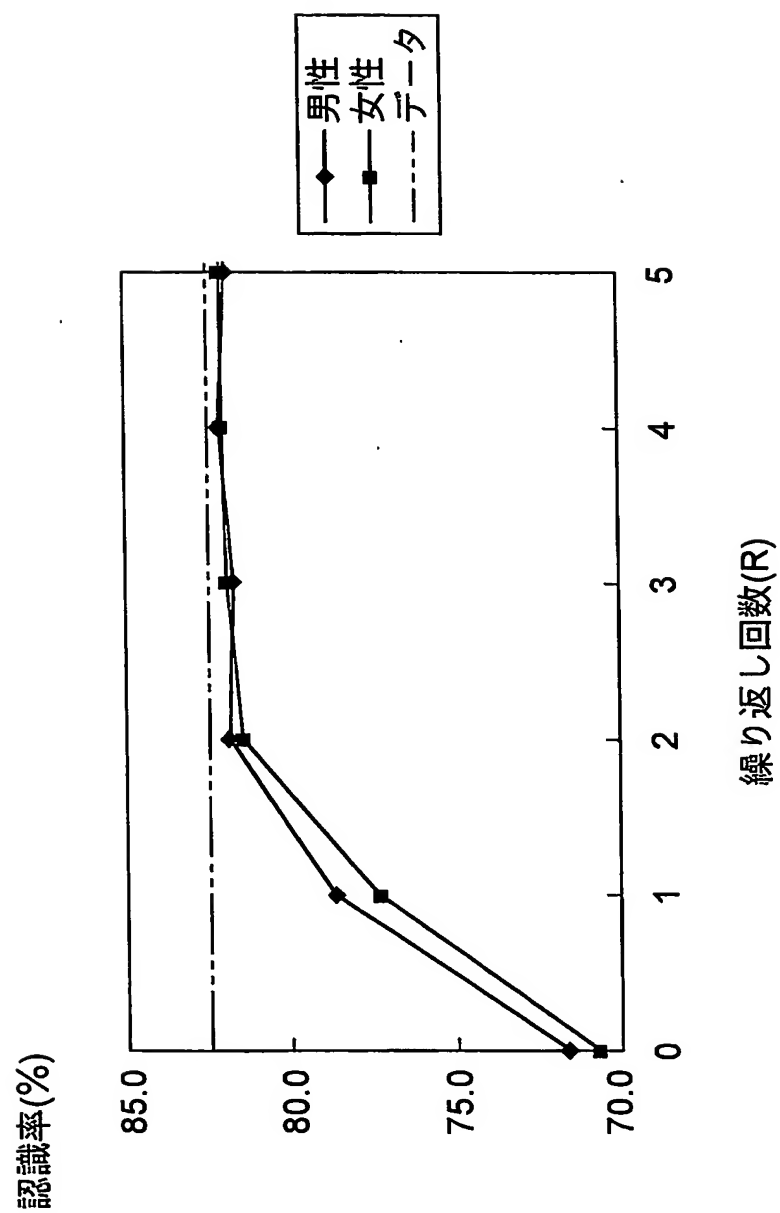


図39

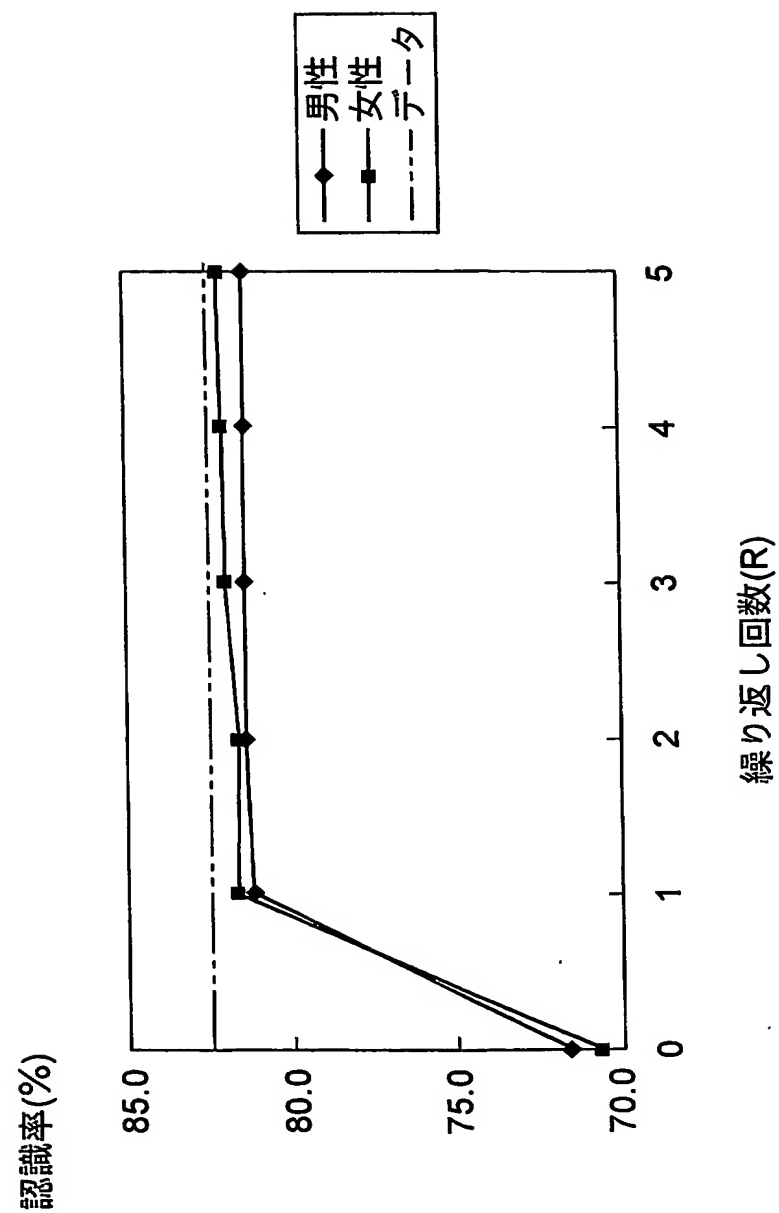


図40

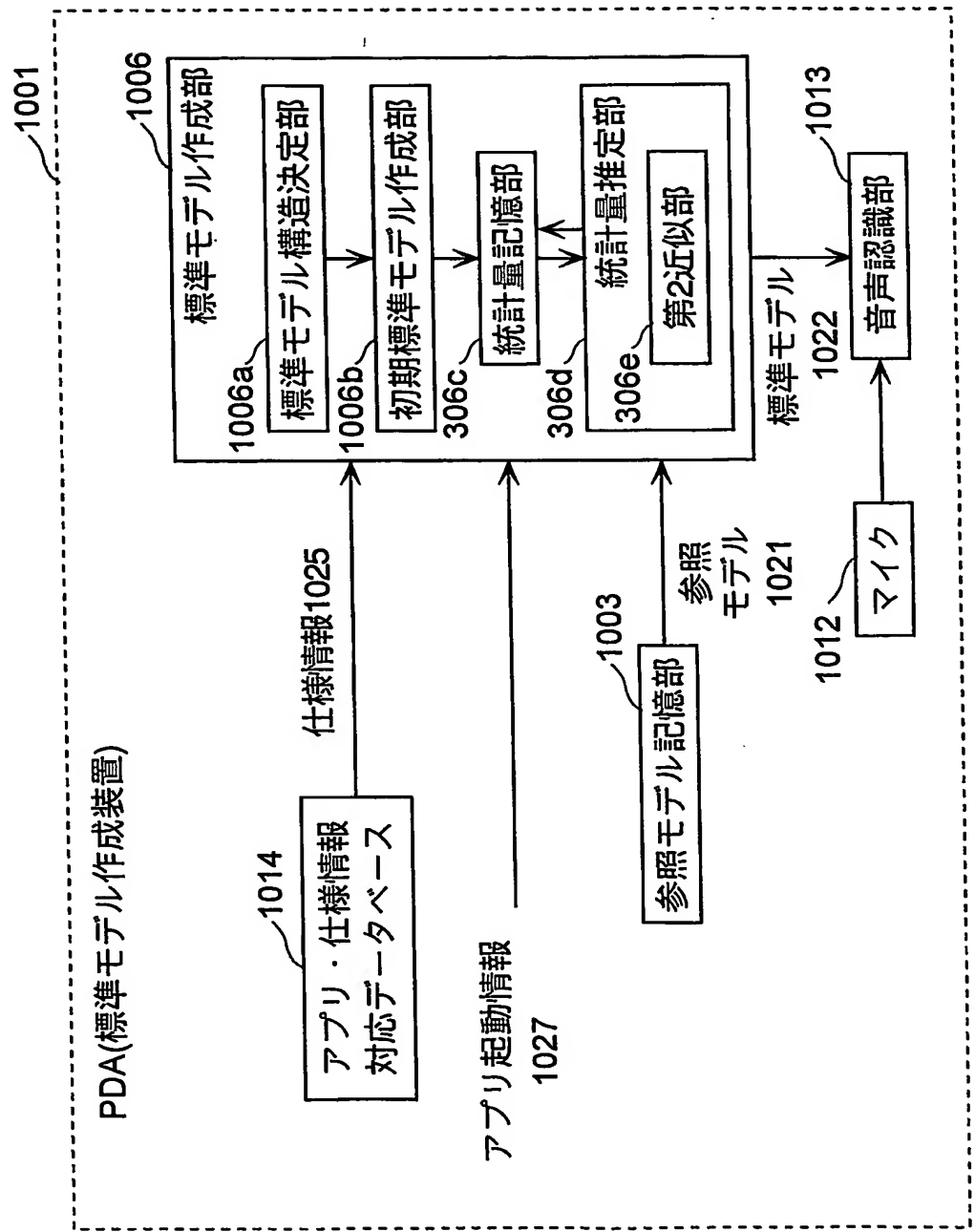


図41

アプリケーション		仕様情報
ID	名前	
1	ゲームA	混合分布数3
2	ゲームB	混合分布数5
3	株取引	混合分布数126
4	テレビのリモコン	混合分布数5
5	翻訳	混合分布数64

図42

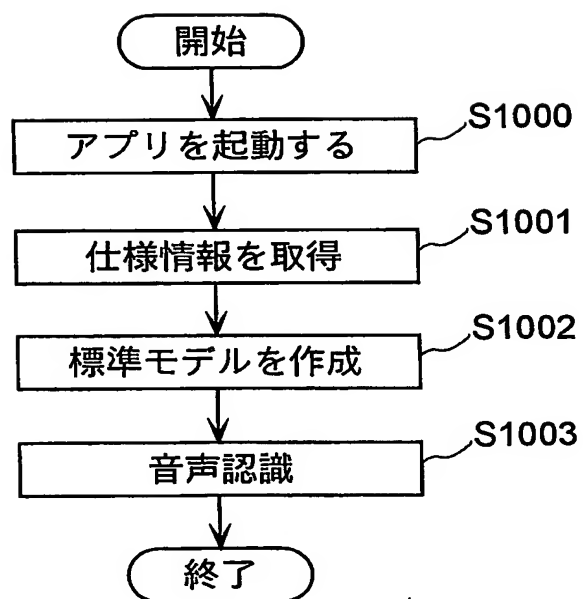


図43

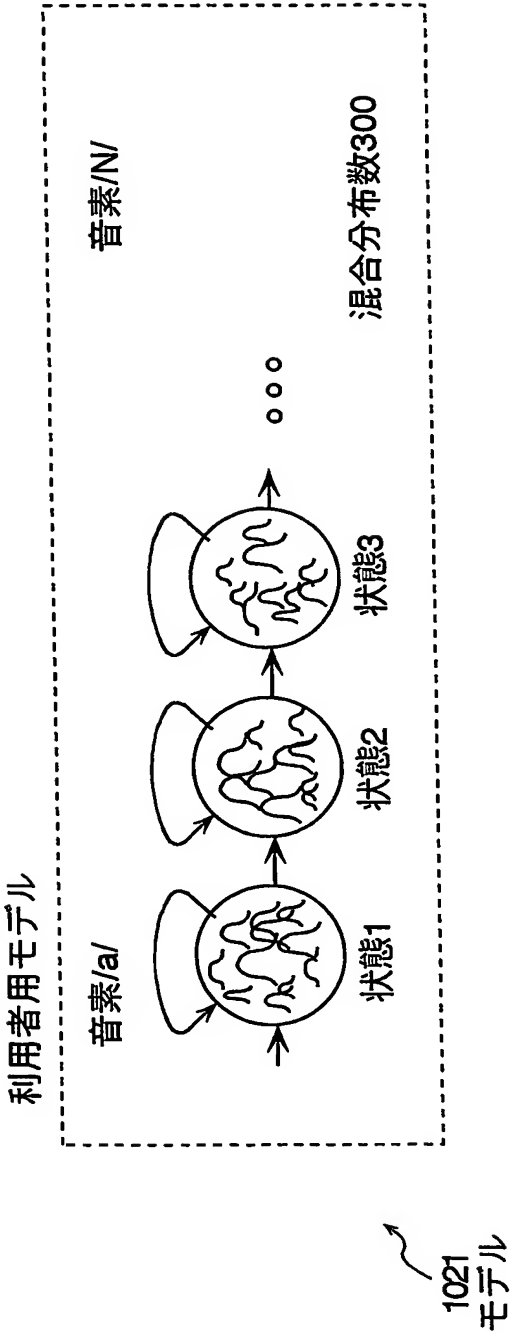


図44

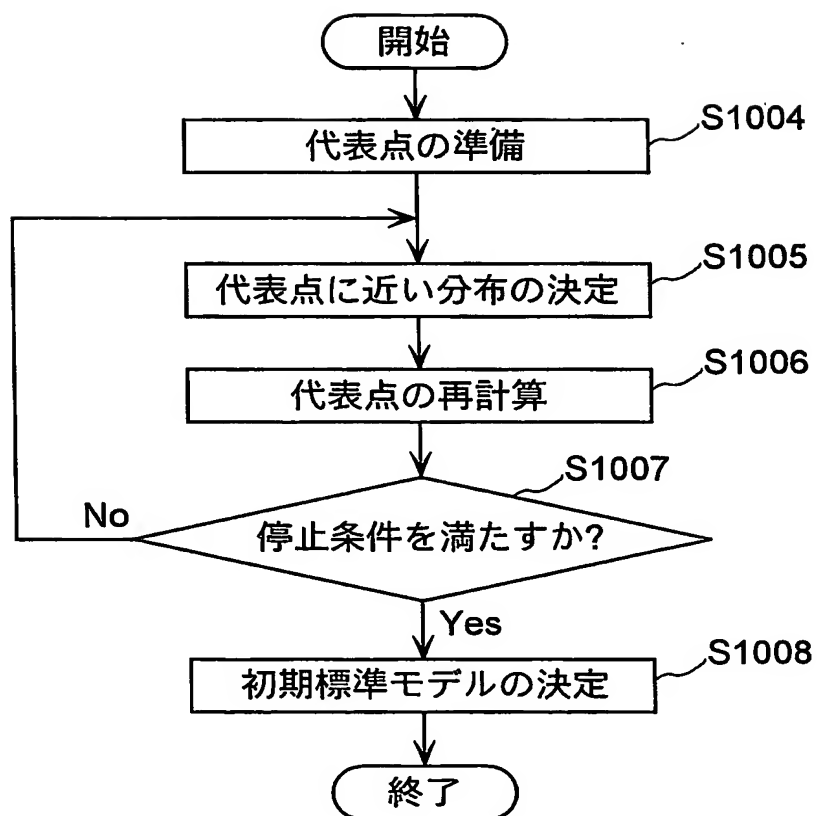


図45

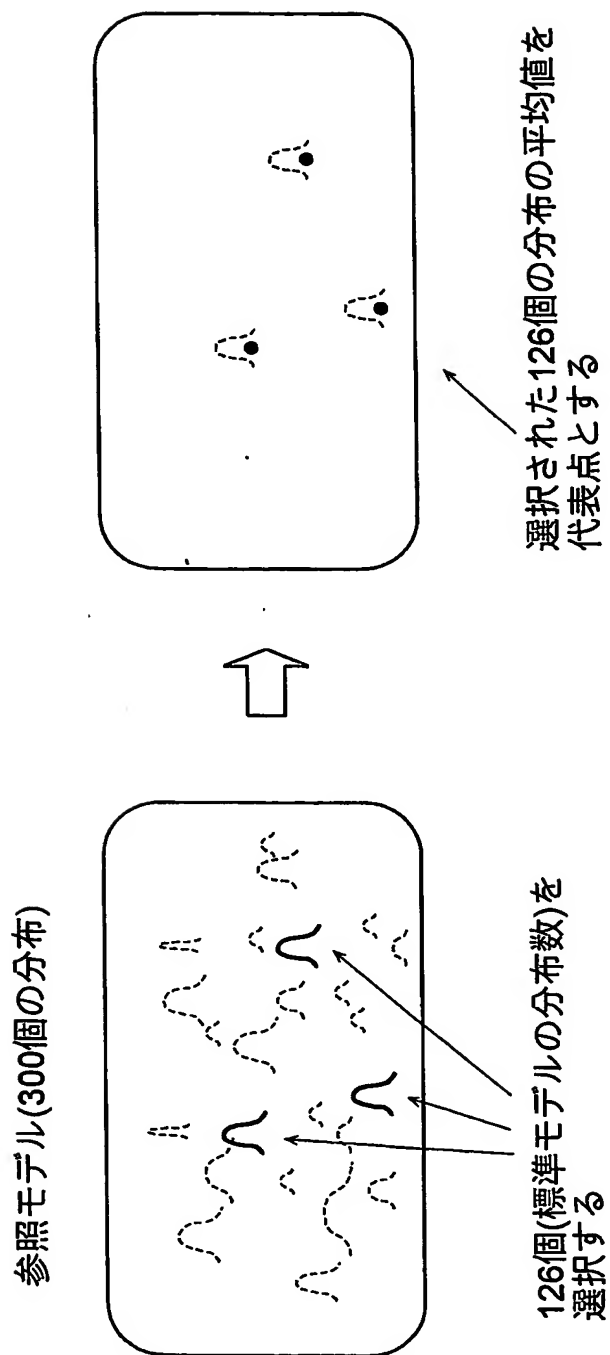


図46

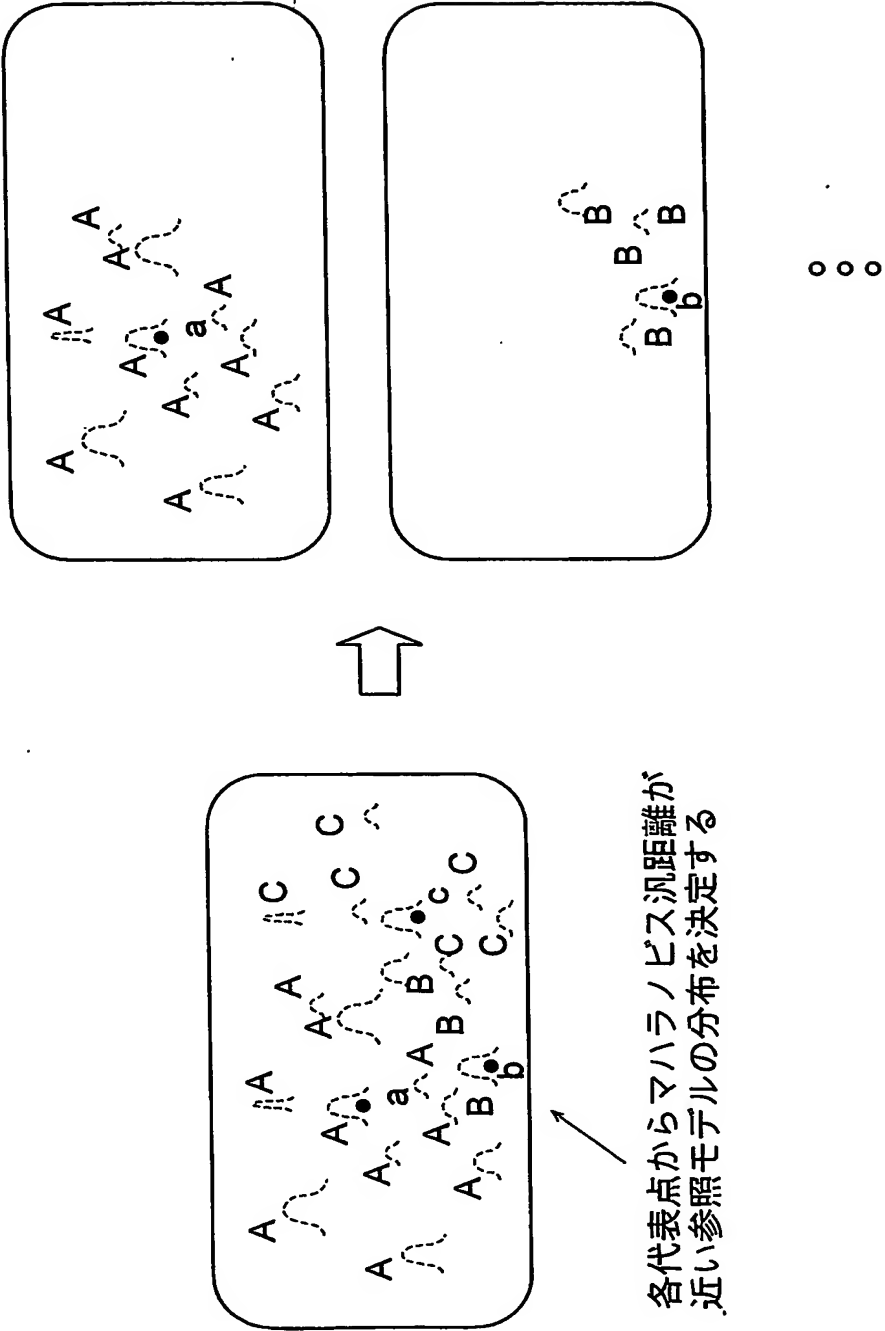


図47

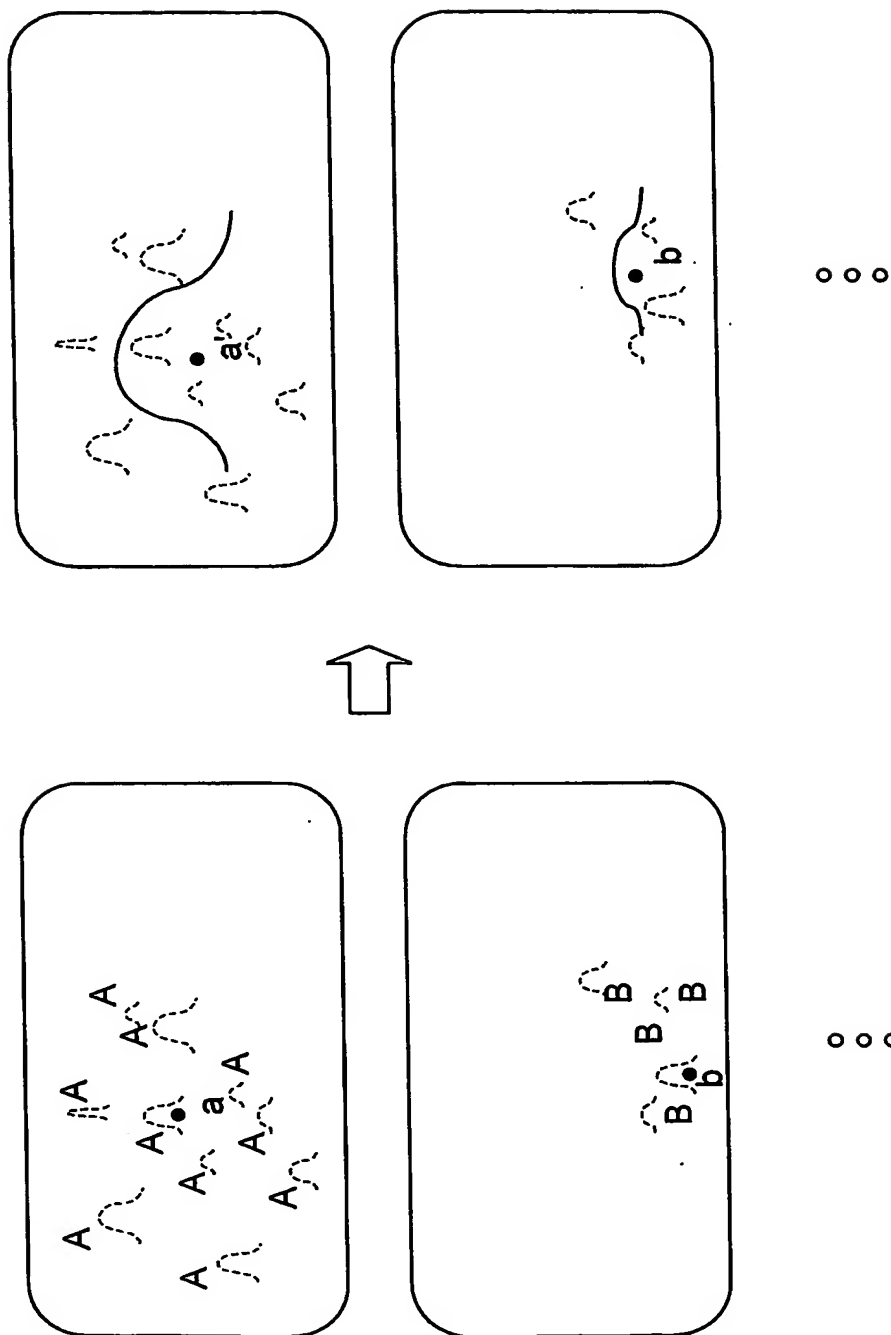


図48

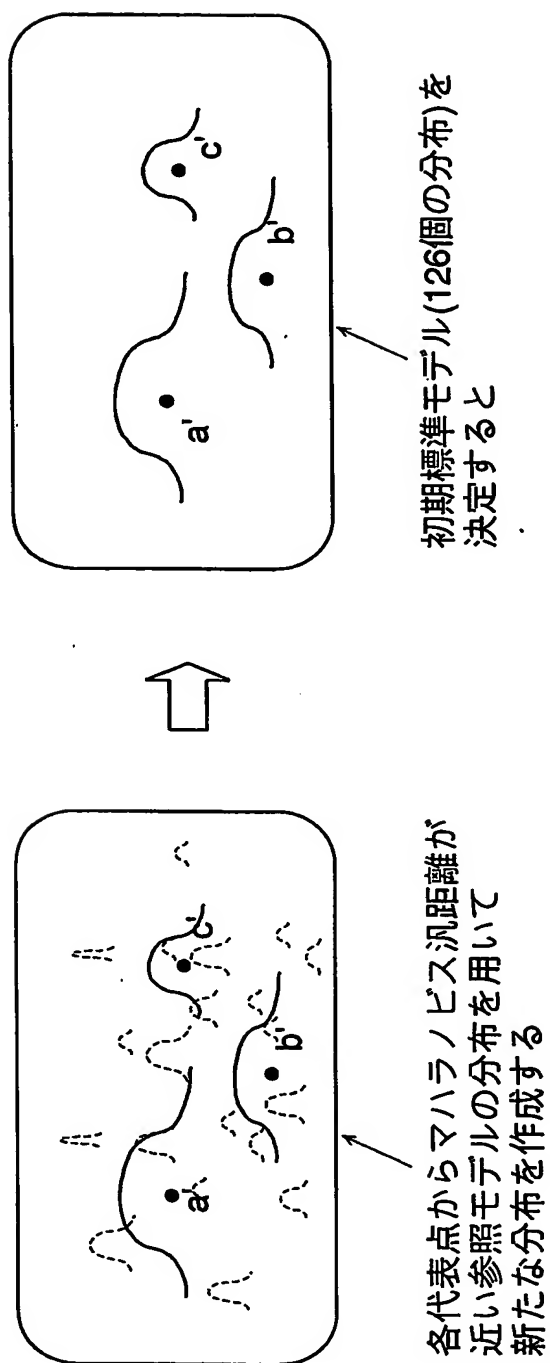


図49

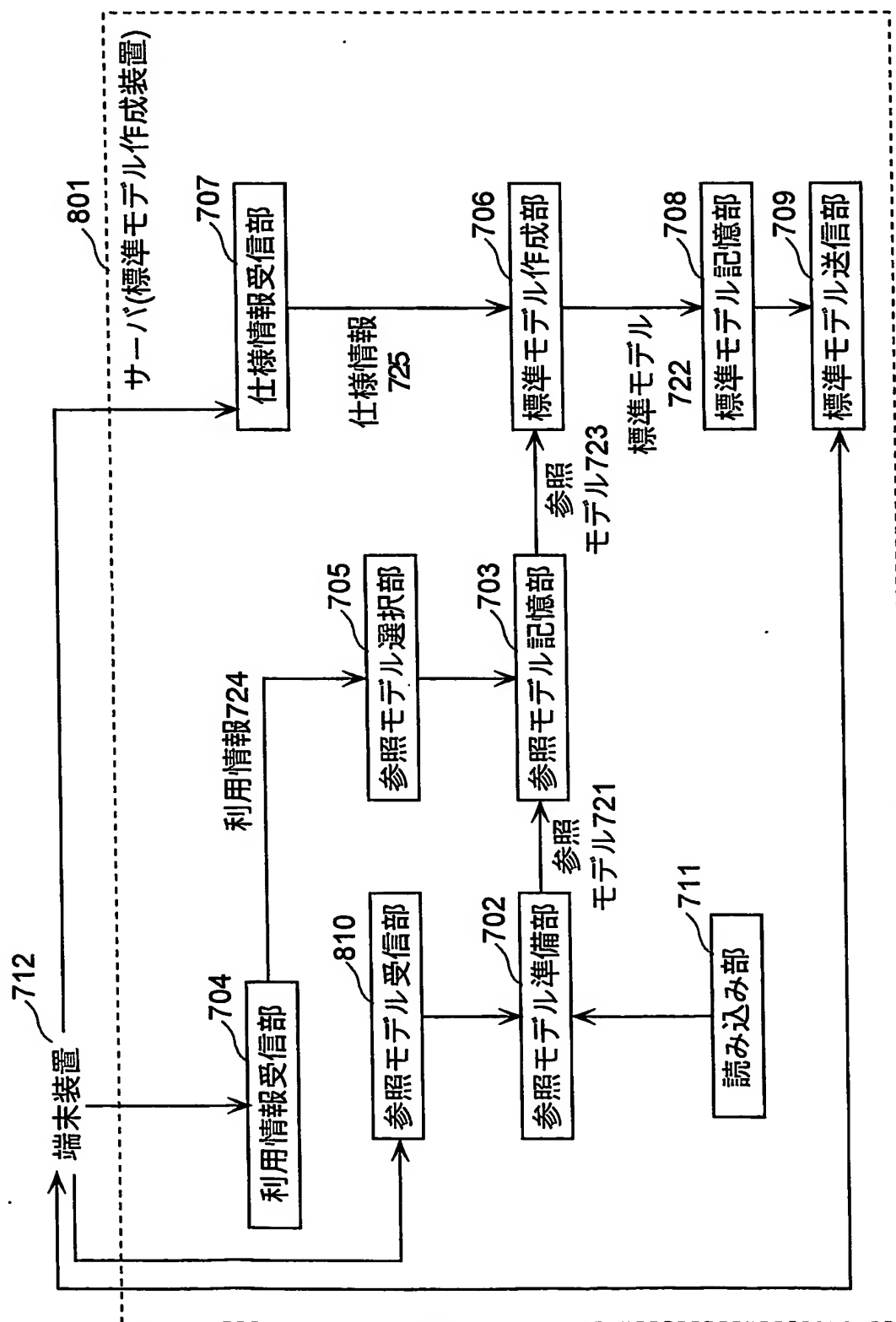


図50

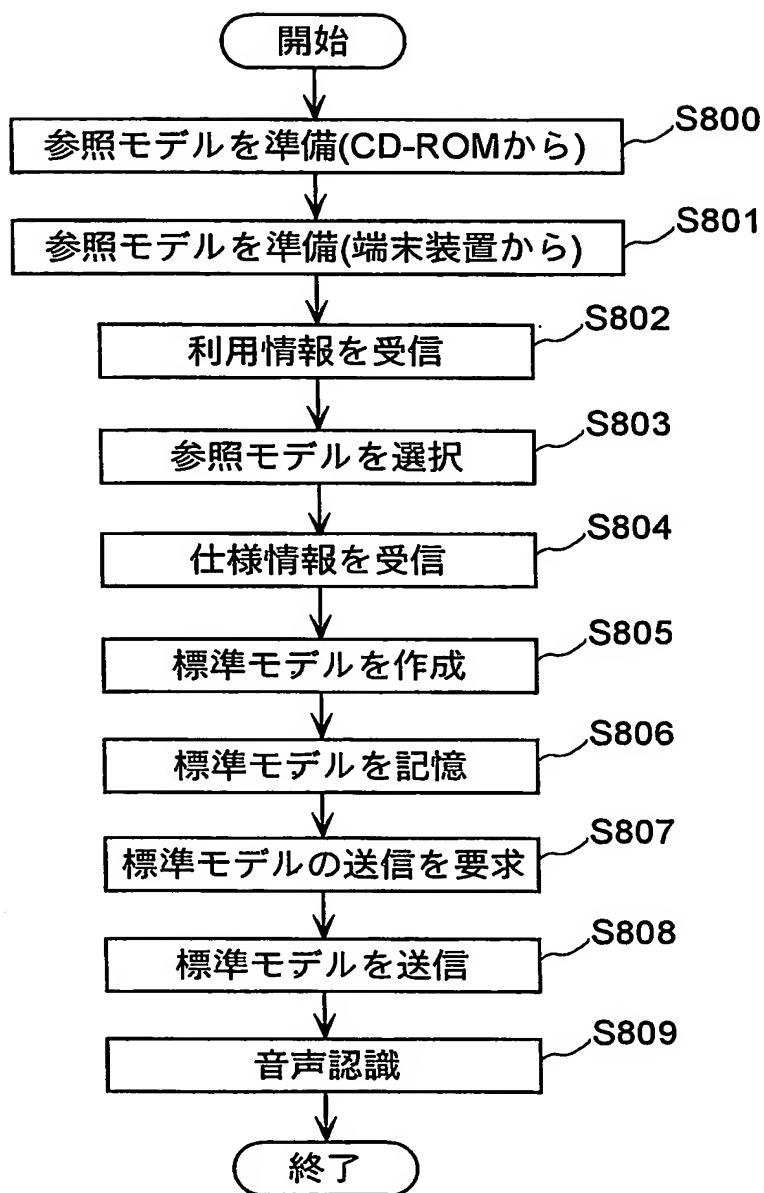


図51

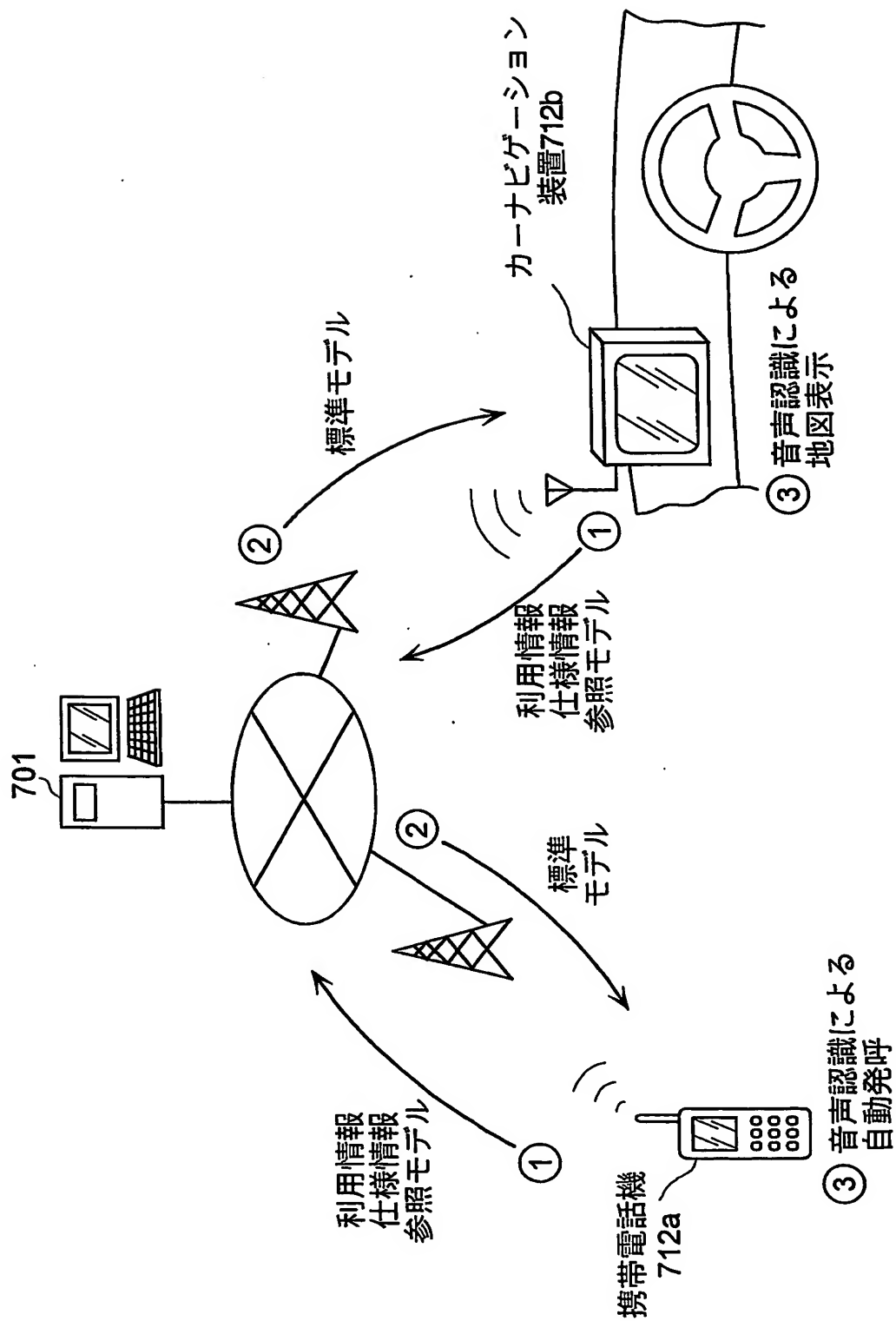


図52

クラスID・初期標準モデル・参照モデル対応表

クラスID	初期標準モデル	参照モデル
8A	初期標準モデル8A	参照モデル8AA
		参照モデル8AB
		参照モデル8AC
		⋮
		参照モデル8AZ
⋮	⋮	⋮
64Z	初期標準モデル64Z	参照モデル64ZA
		参照モデル64ZB
		参照モデル64ZC
		⋮
		参照モデル64ZZ

図53

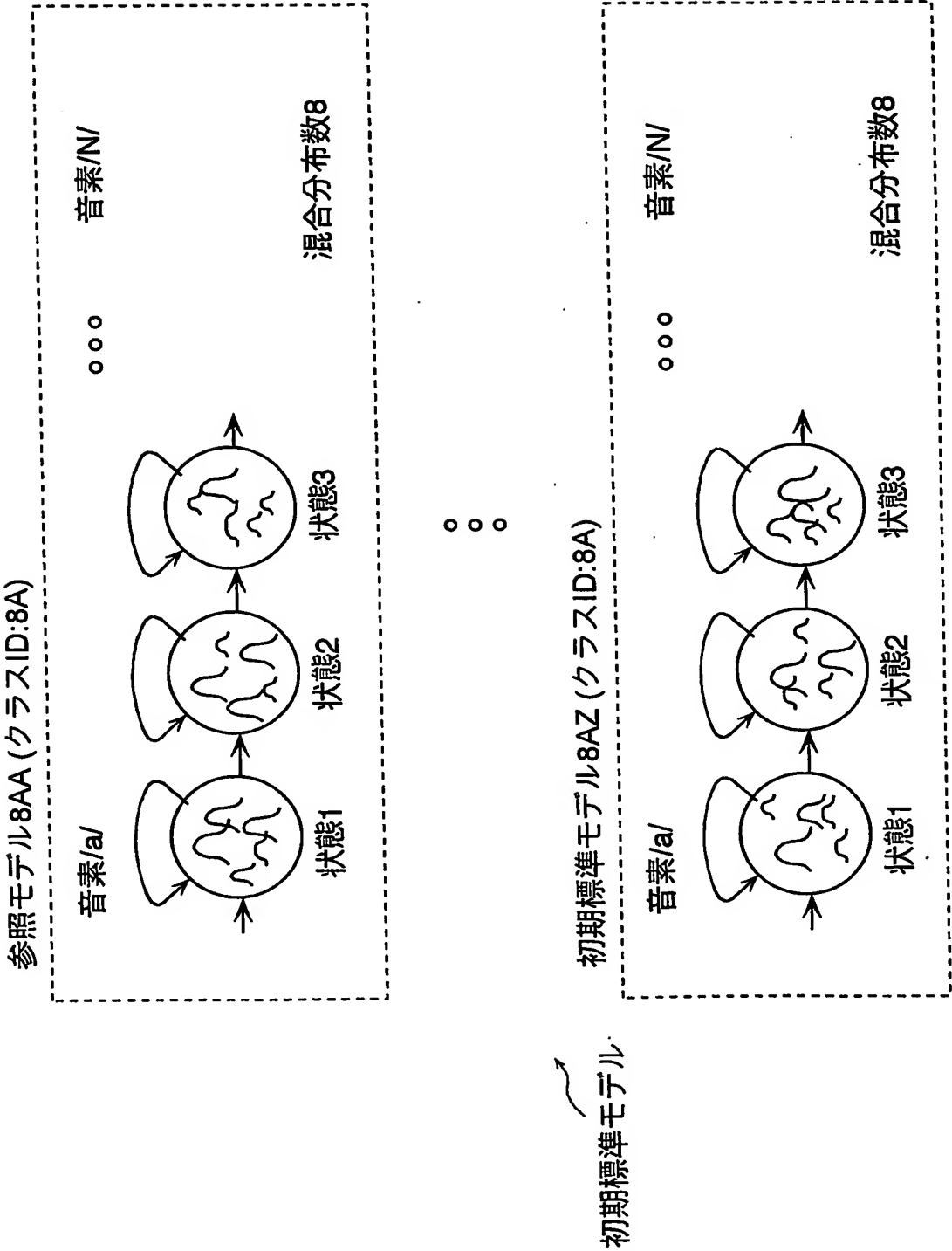


図54

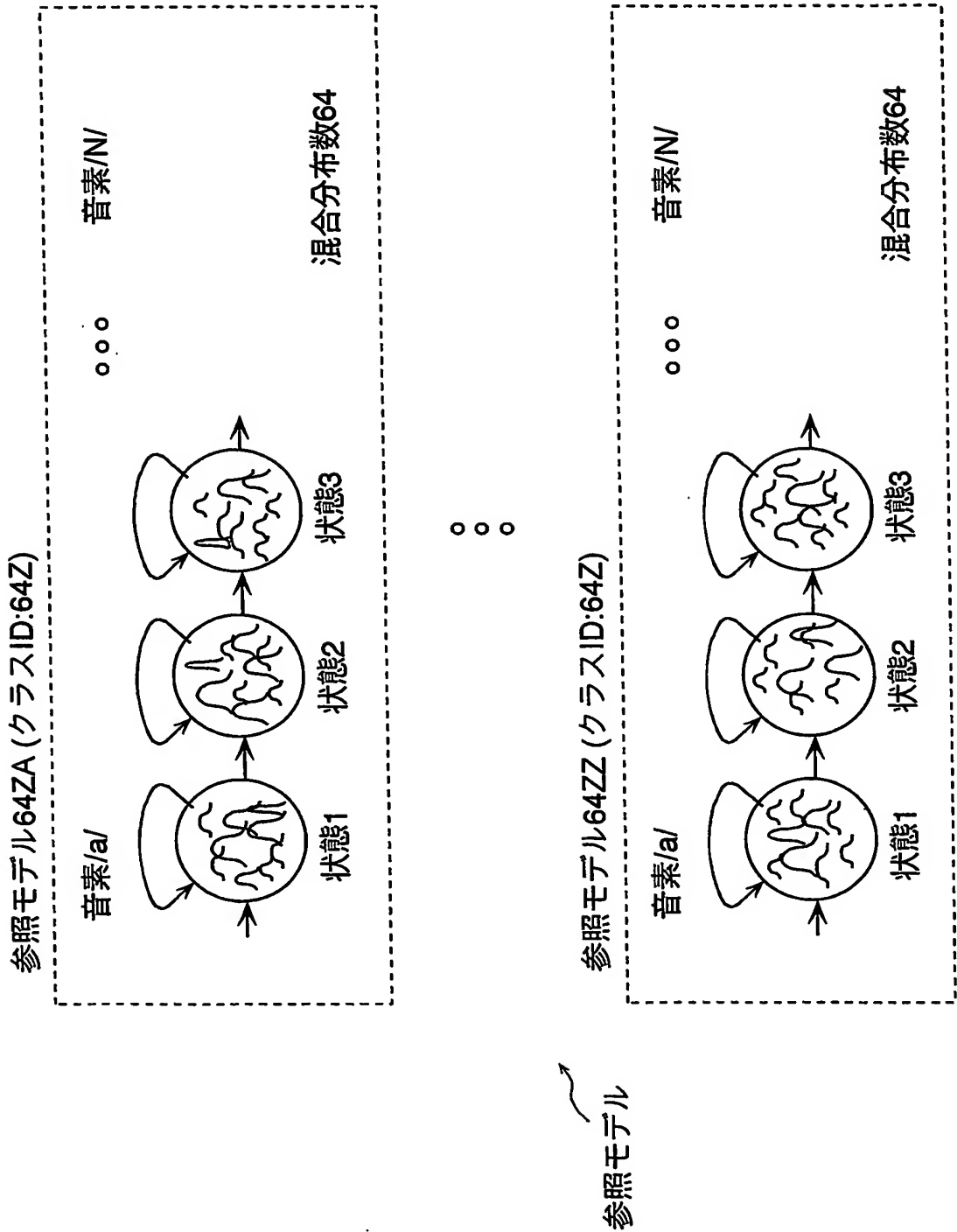


図55

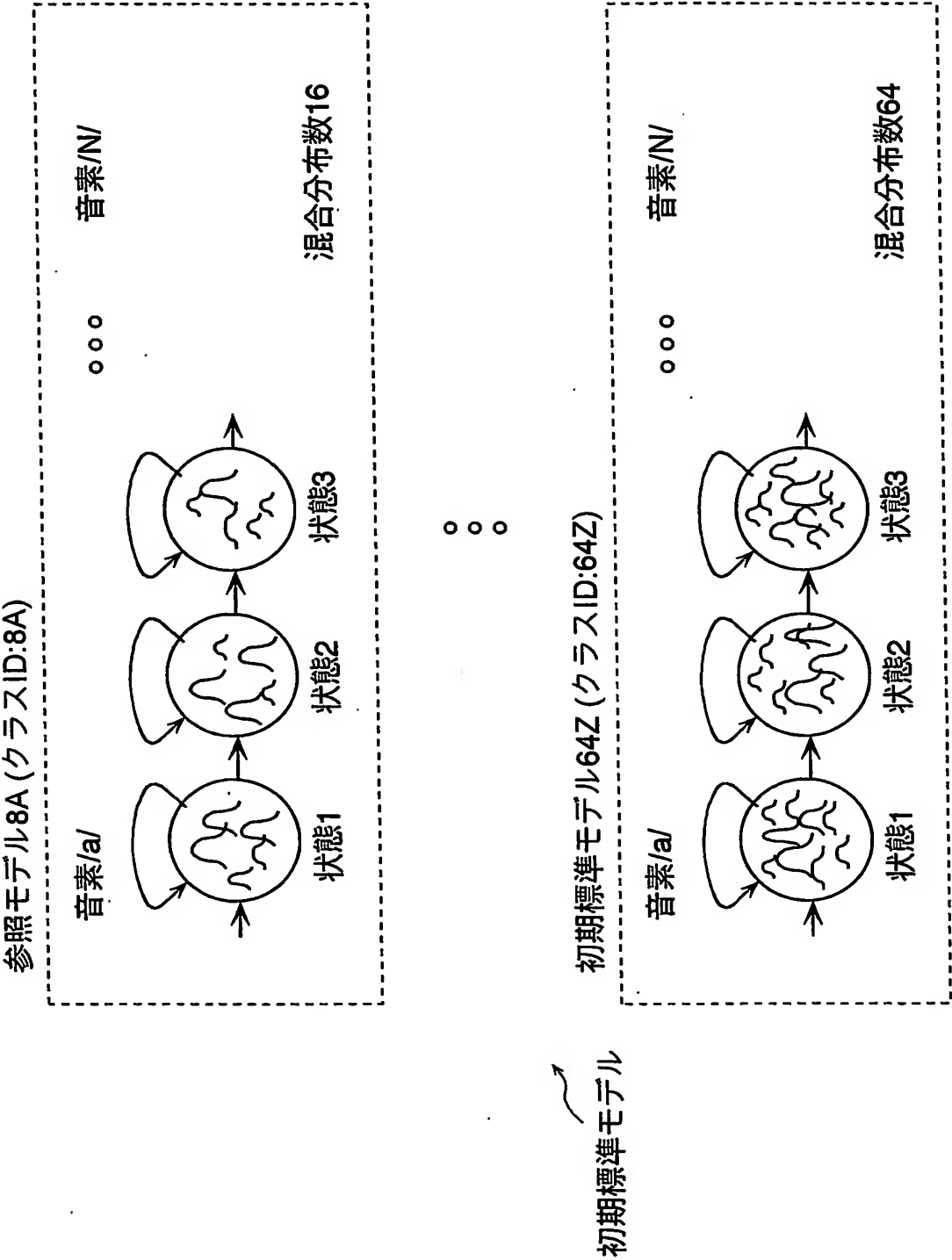


図56

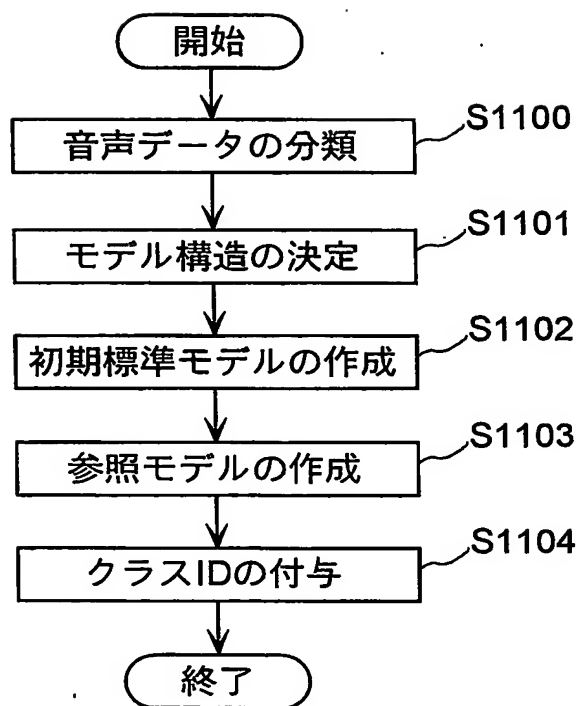


図57

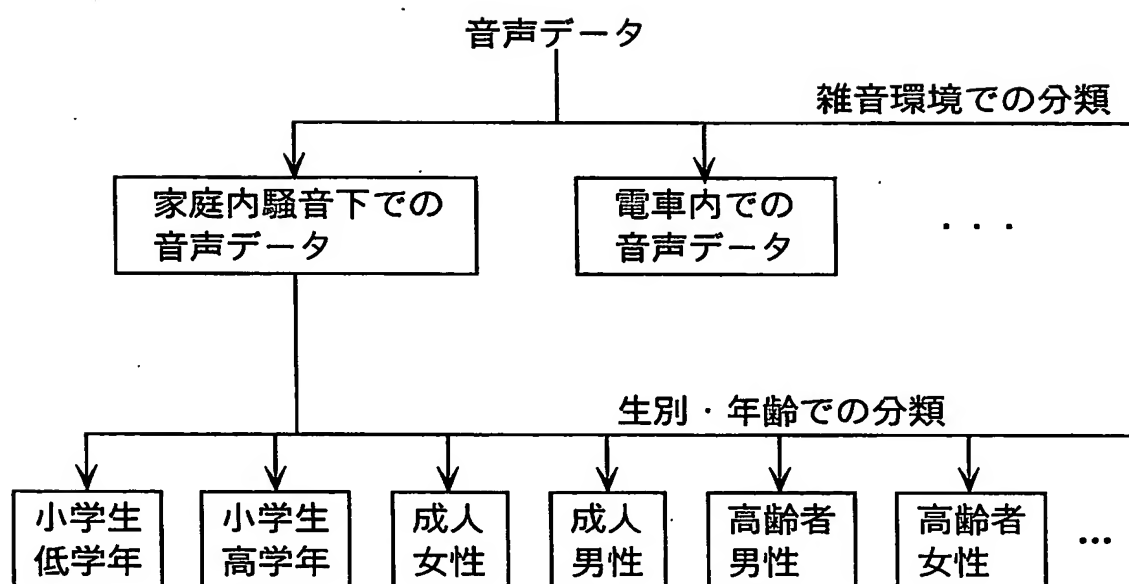


図58

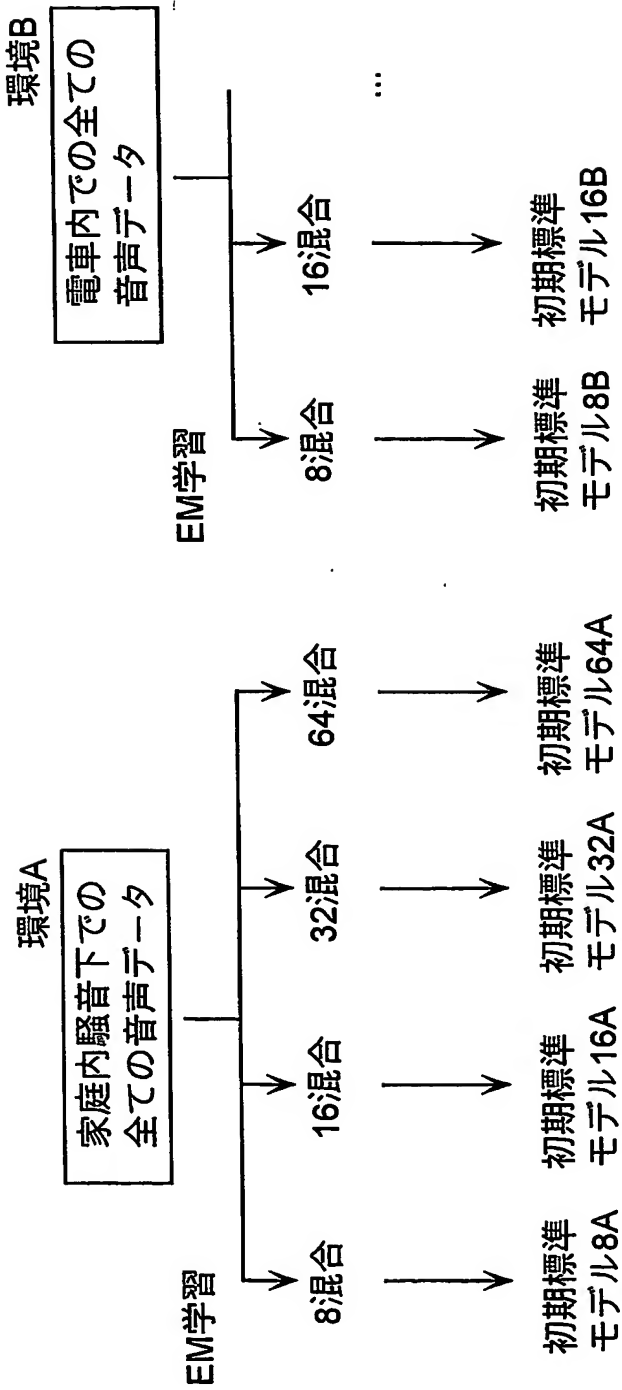


図59

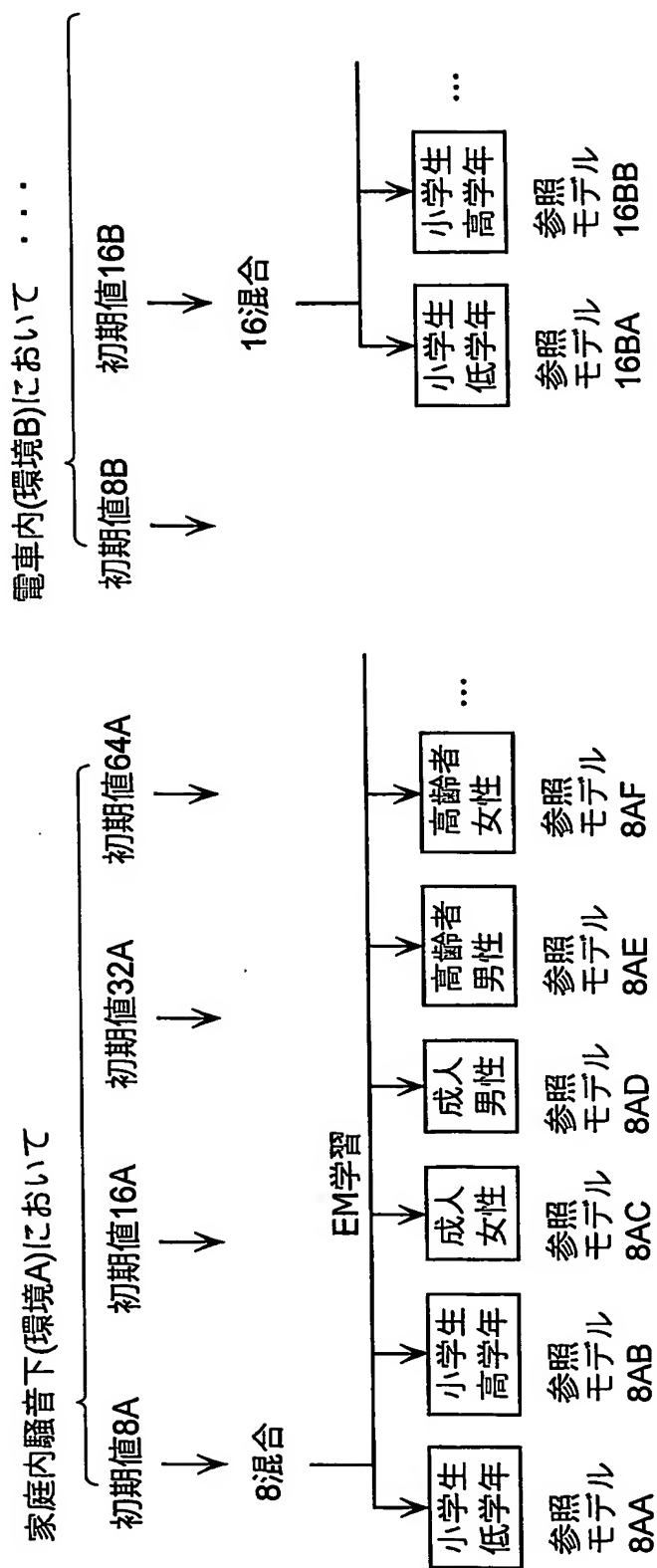


図60

クラスID	初期標準モデル	参照モデル	備考(参照モデルの特長)
8A	初期標準モデル8A	参照モデル8AA 参照モデル8AB 参照モデル8AC ⋮	家庭内騒音, 8混合, 小学校低学年 家庭内騒音, 8混合, 小学校高学年 家庭内騒音, 8混合, 成人女性 ⋮
16A	初期標準モデル16A	参照モデル16AA 参照モデル16AB 参照モデル16AC ⋮	家庭内騒音, 16混合, 小学校低学年 家庭内騒音, 16混合, 小学校高学年 家庭内騒音, 16混合, 成人女性 ⋮
⋮	⋮	⋮	⋮
64B	初期標準モデル64B	参照モデル64BA 参照モデル64BB 参照モデル64BC ⋮	電車内, 64混合, 小学校低学年 電車内, 64混合, 小学校高学年 電車内, 64混合, 成人女性 ⋮

図61

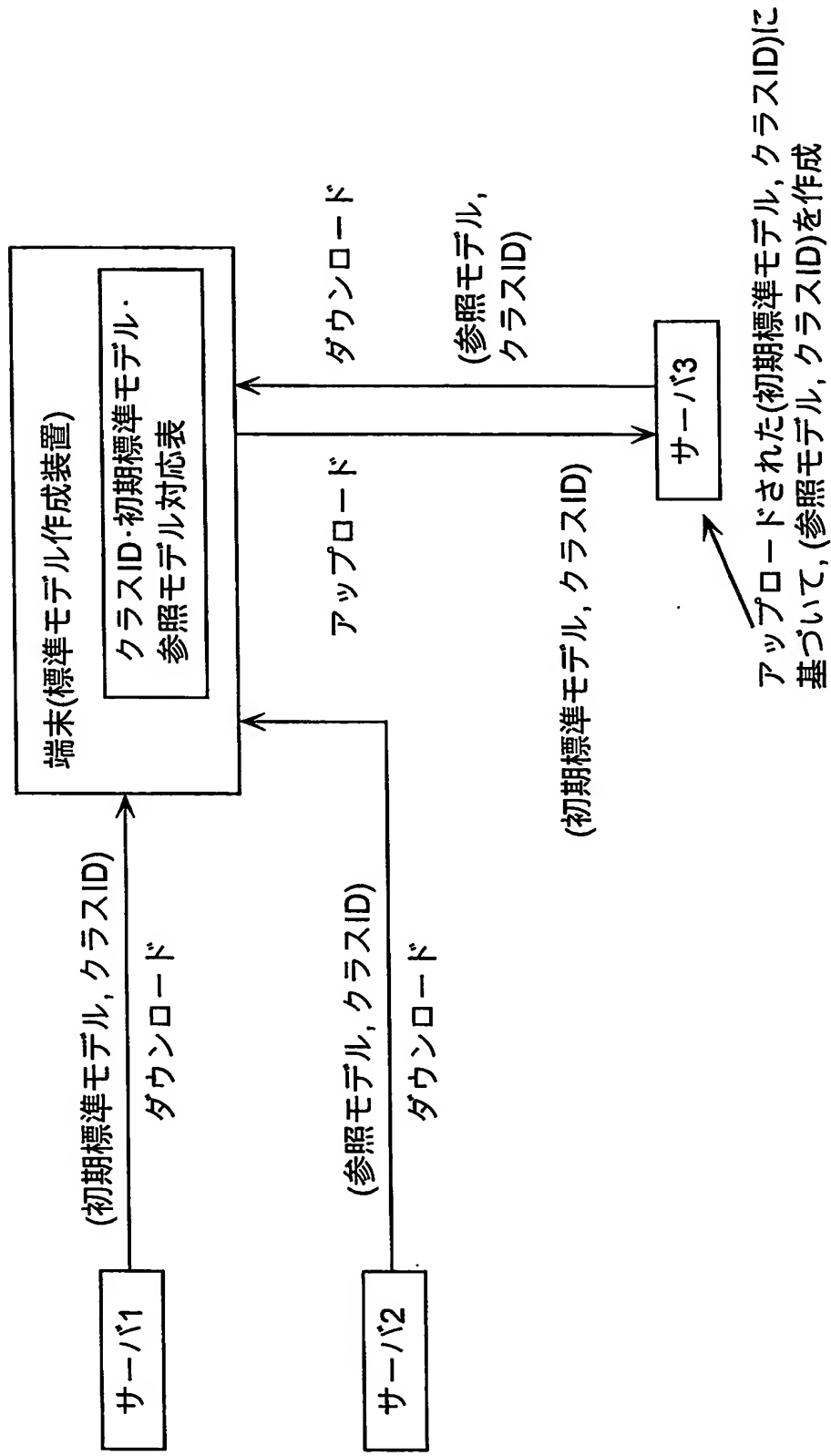


図62

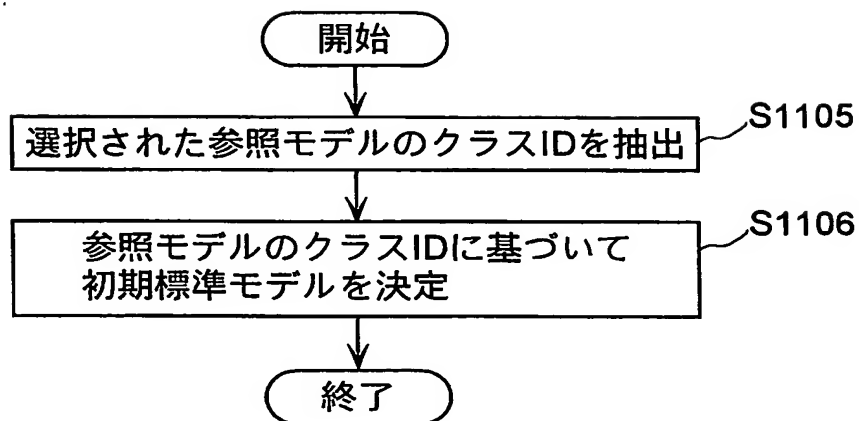


図63

選択された参照モデル	クラスID
参照モデル8AA	8A
参照モデル16AA	16A
参照モデル16AB	16A
参照モデル16AC	16A
参照モデル16BA	16B
参照モデル64BA	64B

図64

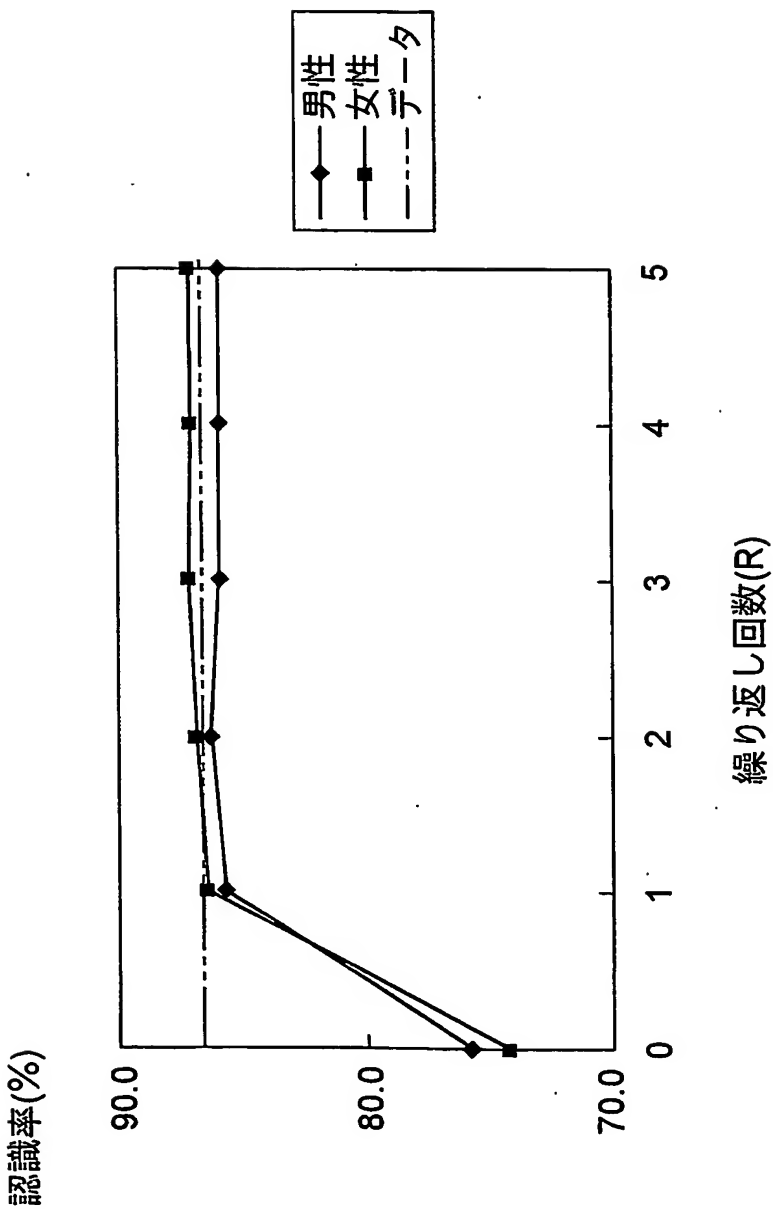
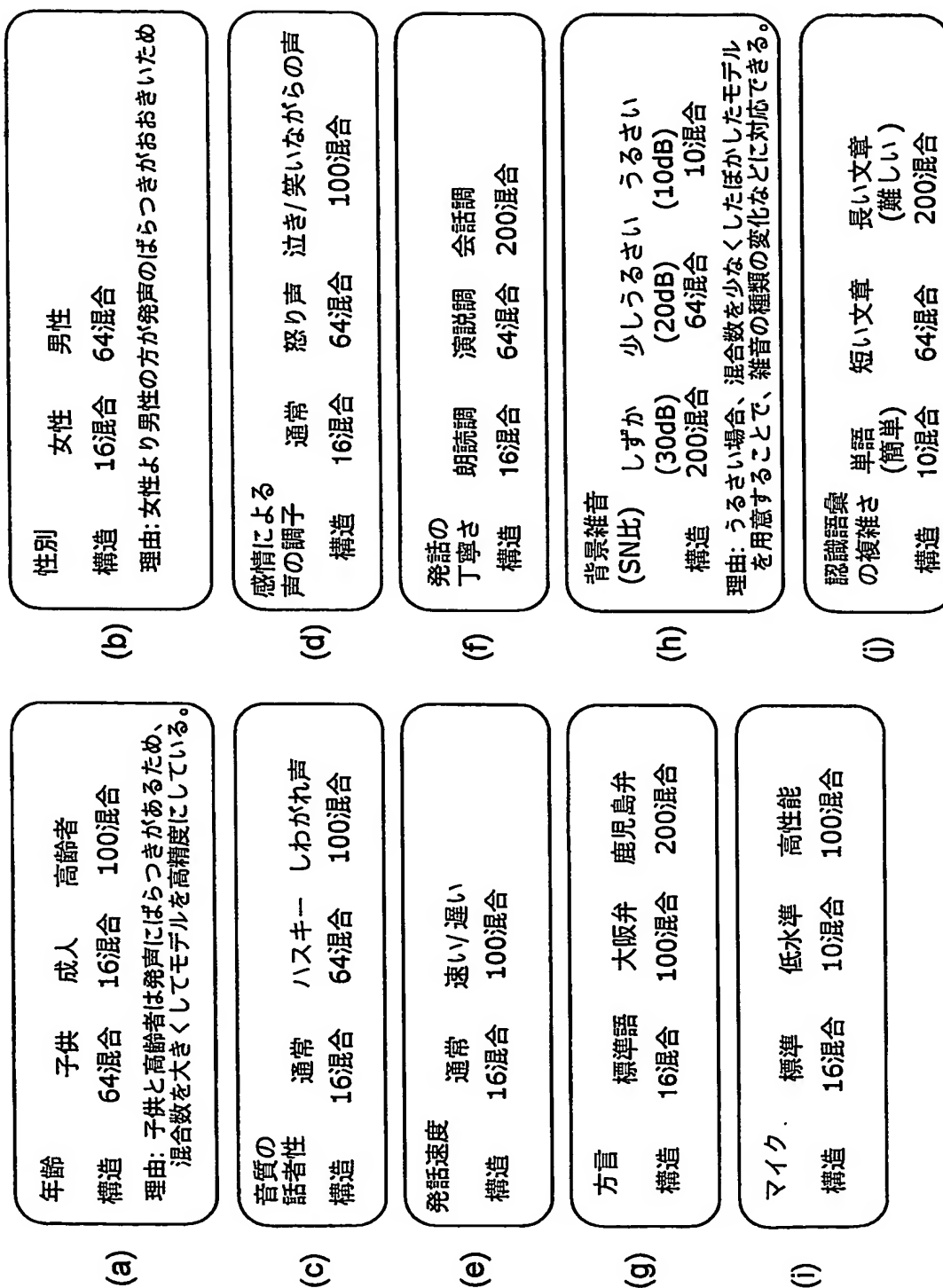


図 5



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/14626

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G10L15/06, G06K9/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl⁷ G10L15/06, G06K9/68

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1926-1995 Toroku Jitsuyo Shinan Koho 1994-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1996-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST FILE(JOIS), IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	YOSHIZAWA, KANO, "Saiyu Suitei ni Motozuku Model Togo Gakushuho", The Acoustical Society of Japan (ASJ) Gakkai 2003 Nen Shuki Kenkyu Happyokai Koen Ronbunshu I, 17 September, 2003 (17.09.03), 3-6-2, pages 105 to 106	1-25
A	YOSHIZAWA et al., "Unsupervised speaker adaptation based on sufficient HMM statistics of selected speakers", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'01), 07 May, 2001 (07.05.01), Vol.1, pages 341 to 344	1-25
A	JP 11-143486 A (Fuji Xerox Co., Ltd.), 28 May, 1999 (28.05.99), Full text; all drawings (Family: none)	1-25

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
06 January, 2004 (06.01.04)

Date of mailing of the international search report
20 January, 2004 (20.01.04)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/14626

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 3251005 B2 (NEC Corp.), 16 November, 2001 (16.11.01), Full text; all drawings (Family: none)	1-25
A	JP 7-69711 B2 (Kabushiki Kaisha ATR Jido Honyaku Denwa Kenkyusho), 31 July, 1995 (31.07.95), Full text; all drawings (Family: none)	1-25
A	JP 2002-236494 A (Denso Corp.), 23 August, 2002 (23.08.02), Full text; all drawings (Family: none)	1-25

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl¹ G10L15/06, G06K9/68

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl¹ G10L15/06, G06K9/68

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926~1995年

日本国公開実用新案公報 1971~2004年

日本国登録実用新案公報 1994~2004年

日本国実用新案登録公報 1996~2004年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS), IEEE Explore

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
P X	芳澤, 鹿野, 「最尤推定に基づくモデル統合学習法」, 日本音響学会2003年秋季研究発表会講演論文集 I, 2003.09.17, 3-6-2, Pages 105-106	1-25
A	Yoshizawa et al, "Unsupervised speaker adaptation based on sufficient HMM statistics of selected speakers", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '01), 2001.05.07, Volume 1, Pages 341-344	1-25

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

06.01.04

国際調査報告の発送日

20.1.2004

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

樫本 剛



5C

9379

電話番号 03-3581-1101 内線 3541

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P 11-143486 A (富士ゼロックス株式会社) 1999.05.28, 全文, 全図 (ファミリーなし)	1-25
A	J P 3251005 B2 (日本電気株式会社) 2001.11.16, 全文, 全図 (ファミリーなし)	1-25
A	J P 7-69711 B2 (株式会社エイ・ティ・アール自動 翻訳電話研究所) 1995.07.31, 全文, 全図 (ファミリーなし)	1-25
A	J P 2002-236494 A (株式会社デンソー) 2002.08.23, 全文, 全図 (ファミリーなし)	1-25